



THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Molecular Investigation of Australian Termites and their Gut Symbionts

Nurdyana Abdul Rahman

BBiotech (Hons), University of Queensland

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2015

School of Chemistry and Molecular Biosciences

Abstract

Termites are one of the most abundant and ecologically important eusocial insects in tropical and subtropical regions. Their success as lignocellulose decomposers is a result of a mutualistic relationship with their gut microbiota. Termites have evolved from wood-feeding cockroaches (lower termites) and expanded their dietary scope to soil, herbivore dung, grass and litter (higher termites). The introduction of high-throughput culture-independent molecular techniques has reinvigorated efforts to understand the termite gut microbiome and its involvement in symbiotic digestion. Yet, there remain significant unanswered or poorly answered questions regarding termite gut microbiome ecology and evolution such as the relative effect of diet vs vertical inheritance on shaping gut communities, the resilience of these communities under changing dietary regimes, the function of specific populations and the relative contributions of prokaryotic and eukaryotic symbionts to hydrolysis in lower termites. The aim of this thesis is to address these questions making use of Australia's diverse but understudied termite species.

In Chapter 2, a molecular survey using SSU rRNA amplicon pyrosequencing was conducted on 66 termite gut samples comprising seven higher termite genera and nine lower termite genera. Findings indicated that vertical inheritance is the primary force shaping the termite gut microbiome, with diet playing a more subtle role changing relative abundance of some populations. This suggested that gut community and structure may change in response to dietary changes as a short-term adaptive mechanism. To test this hypothesis, feeding experiments were performed on the polyphagous lower termite species, *Mastotermes darwiniensis*, and gut communities were monitored over time via SSU rRNA profiling, forming the basis of Chapter 3. Small shifts in relative abundance of gut populations were noted with compositionally different feedstocks (e.g. wood to grass) supporting the original hypothesis, but greater shifts are likely due to response to stress as an effect of smaller colony size. However, only small differences were noted in corresponding gut protein profiles, suggesting that gut function was maintained even though community composition altered. In Chapter 4, whole gut DNA samples of two lower (*Mastotermes* and *Porotermes*) and two higher (*Nasutitermes* and *Microcerotermes*) termite genera were shotgun sequenced. Gene-centric analysis of the shotgun data was performed to determine community-level functional similarities and differences. Despite conspicuous differences in community structure between these four wood-feeding genera (Chapter 2), they had similar gene family abundance profiles suggesting functional convergence in these communities. Differential coverage binning was also attempted to recover population genomes from the metagenomic datasets, but with limited success due to termite host DNA compromising the assemblies. However, three and one substantially complete population genomes belonging to the Fibrobacteres and TG3 phyla, respectively, were recovered from the two

higher termite genera, *Microcerotermes* and *Nasutitermes*. Currently only one sequenced isolate is publicly available for the Fibrobacteres (*Fibrobacter succinogenes*) and one for TG3 (*Chitinivibrio alkaliphilus*). In Chapter 5, a comparative genomics analysis of the Fibrobacteres/TG3 phyla was conducted using the four termite gut population genomes together with three from a cellulose-fed anaerobic digester, one from sheep rumen and the two isolate reference genomes. Genome-based phylogeny indicated a robust relationship between Fibrobacteres and TG3, thus we propose reclassifying TG3 as a class within the Fibrobacteres phylum. Polymer hydrolysing genes were found to be over-represented in all 10 genomes suggesting that this is a unifying characteristic of the Fibrobacteres, although additional ecosystems should be investigated to confirm this inference. Historically, the Fibrobacteres were thought to be non-motile based on extrapolation from *F. succinogenes*, however, our analysis suggests that flagella-based motility is an ancestral and widespread trait in this phylum and has been recently lost in *F. succinogenes* and related genera.

The findings of this thesis contribute to the growing body of knowledge on termite gut microbiomes and in particular, Australian species. The Fibrobacteres genomes provide insight into the evolution of this understudied and underrepresented phylum, which are common constituents of anoxic fibrolytic communities.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Published peer-reviewed papers

Abdul Rahman, N., Parks, D.H., Willner, D.L., Engelbrektson, A.L., Goffredi, S.K., Warnecke, F., Scheffrahn, R. H. and Hugenholtz, P. 2015. A molecular survey of Australian and North American termite genera indicates that vertical inheritance is the primary force shaping termite gut microbiomes. *Microbiome* **3**: 5.

Abdul Rahman, N., Parks, D.H., Vanwonderghem, I., Morrison, M., Tyson, G. W., Hugenholtz, P., 2015. A phylogenomic analysis of the bacterial phylum Fibrobacteres. *Frontiers in Microbiology* **6**: 1469.

Publications included in this thesis

Abdul Rahman, N., Parks, D.H., Willner, D.L., Engelbrektson, A.L., Goffredi, S.K., Warnecke, F., Scheffrahn, R. H. and Hugenholtz, P. 2015. A molecular survey of Australian and North American termite genera indicates that vertical inheritance is the primary force shaping termite gut microbiomes. *Microbiome* **3**: 5 – incorporated as Chapter 2.

Contributor	Statement of contribution
Abdul Rahman, N. (Candidate)	Designed experiments (35%) Performed experiment (40%) Analysed data (55%) Wrote and edited paper (55%)
Parks, D.H.	Analysed data (20%)
Willner, D.L.	Analysed data (5%)
Engelbrektson, A.L.	Performed experiment (30%)
Goffredi, S.K.	Performed experiment (10%)
Warnecke, F.	Designed experiments (30%) Performed experiment (10%)
Scheffrahn, R. H.	Performed experiment (10%)
Hugenholtz, P.	Designed experiments (35%) Analysed data (20%) Wrote and edited paper (45%)

Abdul Rahman, N., Parks, D.H., Vanwonderghem, I., Morrison, M., Tyson, G. W., Hugenholtz, P., 2015. A phylogenomic analysis of the bacterial phylum Fibrobacteres. *Frontiers in Microbiology* – incorporated as Chapter 5.

Contributor	Statement of contribution
Abdul Rahman, N. (Candidate)	Designed experiments (30%) Performed experiment (55%) Analysed data (55%) Wrote and edited paper (50%)
Parks, D.H.	Analysed data (20%) Wrote and edited paper (5%)
Vanwonderghem, I.	Performed experiment (45%) Analysed data (5%) Wrote and edited paper (2%)
Morrison, M.	Designed experiments (5%) Analysed data (5%) Wrote and edited paper (5%)
Tyson, G. W.	Designed experiments (5%) Wrote and edited paper (8%)
Hugenholtz, P.	Designed experiments (60%) Analysed data (15%) Wrote and edited paper (30%)

Contributions by others to the thesis

Chapter 3 – Effect of diet on microbial community composition in the gut of *Mastotermes darwiniensis*

Study design and chapter review: Maria Chuvochina and Philip Hugenholtz

Chapter 4 – Metagenome-based analysis of gut communities in lower and higher termites

Study design and chapter review: Philip Hugenholtz

Statement of parts of the thesis submitted to qualify for the award of another degree

None

Acknowledgements

Alhamdulillah, all praises to Allah for the strengths and His blessings in completing this thesis. Without the help of the people, to whom I wish to express my gratitude, this thesis could not be completed. First and foremost, I'd like to thank my PhD advisors, Professors Philip Hugenholtz, Gene Tyson and David Merritt, for their consistent guidance, unparalleled knowledge, tutelage, and encouragement during my candidature. Phil is the coolest advisor and one of the smartest people I know. His enthusiasm and dedication towards his work has greatly influenced me in becoming a better researcher. Gene has been supportive and has provided insightful comments about my research.

I'm also grateful to mentors, Dr. Dana Willner, Dr. Donovan Parks and Dr. Maria Chuvochina for their scientific advice, knowledge, and many insightful discussions and suggestions. Dana was my primary resource for getting my bioinformatics questions answered and was instrumental in helping me start my research. I'd also like to thank the members of my PhD committee, Professors Mark Walker, Mark Morrison, and Phil Bond for their helpful advice and suggestions during my milestones.

I'd especially like to thank my best friend, Serene Low for her constant support and putting up with all my rants. Thank you to the friends that I've made through the course of my studies at UQ who have kept me sane through this journey. Lendl Tan for his advice and great company and my awesome housemates, Halimah Mahmud, Dayana Mahiuddin and Akmaliza Ali, thank you for making me feel at home. I'd like to thank Kelvin Goh, Yun Kit Yeoh and Sarah Yuzaidi for reading my thesis chapters. To the ACE mates; Rochelle Soo, Josh Daly, Inka Vanwonterghem, Nancy Lachner, Caitlin Singleton and Fauzi Haroon, thanks for the help and making ACE a great learning place. Also, I'm most fortunate to have good friends like Amir Arshad, Khadijah Samsudin, Rafta Jamal, and Rasul Sulaiman [Type equation here](#).who are always being there for me.

Last, but not least, a special thanks to my family. Words cannot express how grateful I'm to my mother, Sanbiah Nasan and father, Abdul Rahman for the sacrifices that you have made. Thank you for always believing in me and encouraging me to follow my dreams. I'd also like to thank my sister, Syda Rahman, brothers, Rafie and Rashid Rahman and aunt, Sutimah Nasan, for supporting and encouraging me throughout this experience, and to my beloved nieces, Qistina, Qisyah, Qaireen and Quraisya for always cheering me up.

I'm also grateful to the UQ International Postgraduate Research Scholarship and UQ Research Scholarship that financially supported me during my candidature.

Keywords

Termite gut, Australian termite, termite feeding trials, Fibrobacteres, comparative genomics, metabolism, phylogenetics

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 060504, Microbial Ecology, 50%

ANZSRC code: 060309, Phylogeny and Comparative Analysis, 20%

ANZSRC code: 060104, Cell Metabolism, 30%

Fields of Research (FoR) Classification

FoR code: 0605, Microbiology, 70%

FoR code: 0601, Biochemistry and cell biology, 30%

Table of Contents

Chapter 1 Introduction and Literature Review	1
1.1 Insect-microbial interactions	1
1.1.1 Symbiosis	1
1.1.2 Benefits of resident microbiota on the insect host	2
1.2 Termites as a model insect for basic and applied research.....	3
1.2.1 Evolution and diet	3
1.2.2 Australian termites	4
1.2.3 Ecology	6
1.2.4 Diversity and distribution of microorganisms in the termite gut.....	8
1.2.3.1 Symbiotic gut flagellates	9
1.2.3.2 Symbiotic gut bacteria and archaea.....	10
1.3 Culture-independent molecular approaches to study microbial ecosystems.....	10
1.3.1 SSU rRNA-based methods.....	10
1.3.2 Genome-based methods	12
1.3.2.1 Reads quality trimming and assembly.....	14
1.3.2.2 Binning.....	15
1.3.2.3 Gene prediction and annotation.....	15
1.3.2.4 Metabolic reconstruction and comparative analysis	17
1.3.2.5 Metatranscriptomics and metaproteomics.....	18
1.4 Application of culture-independent molecular methods to termite gut communities	19
1.4.1 Discovery of novel bacterial diversity in the termite gut.....	23
1.4.2 Metagenomic analyses of termite gut microbiomes	24
1.4.3 Expressed genes in termite guts.....	25
1.5 Summary of results chapters	26
1.6 References	27
Chapter 2 A molecular survey of Australian and North American termite genera indicates that vertical inheritance is the primary force shaping termite gut microbiomes	42
2.1 Introduction	44
2.2 Methods.....	44
2.2.1 Sample collection and processing.....	44
2.2.2 DNA extraction.....	45
2.2.3 SSU rRNA PCR and amplicon pyrosequencing.....	45
2.2.4 Analysis of SSU rRNA gene sequences	46
2.2.5 Molecular identification of termite host species.....	47
2.2.6 Stable isotope analysis	47
2.2.7 Nucleotide sequence accession numbers	47
2.3 Results	48
2.3.1 Sample collection and host identification.....	48
2.3.2 Gut microbiome profiling	49
2.3.3 Bacterial profiles	50
2.3.4 Archaeal profiles.....	55
2.3.5 Eukaryotic profiles.....	57

2.3.6	Beta-diversity analyses	57
2.4	Discussion	60
2.5	Conclusions	65
2.6	Acknowledgements	66
2.7	References	66
Chapter 3 Effect of diet on microbial community composition in the gut of <i>Mastotermes</i>		
<i>darwiniensis</i>..... 73		
3.1	Introduction	74
3.2	Methods.....	75
3.2.1	Rearing of termites under laboratory conditions.....	75
3.2.1.1	Optimising maintenance conditions for feeding experiments.....	75
3.2.1.2	Feeding experiments.....	76
3.2.2	Microbial community profiling of the gut of <i>M. darwiniensis</i>	77
3.2.3	PCR amplification and amplicon sequencing.....	77
3.2.4	Sequence processing and analysis.....	78
3.2.5	Proteomics.....	79
3.2.5.1	Preparation of crude enzyme extract.....	79
3.2.5.2	Measurement for cellulase activity.....	79
3.2.5.3	Protein electrophoresis and zymogram analysis	80
3.2.5.4	Protein sequencing and identification	81
3.3	Results and Discussion.....	82
3.3.1	Optimisation of rearing termites under laboratory conditions.....	82
3.3.2	Effect of different lignocellulose feedstocks on gut microbiota of <i>Mastotermes</i>	84
3.3.2.1	<i>Mastotermes</i> gut community profiling.....	84
3.3.2.2	Variation of gut microbial profiles of <i>Mastotermes</i> using different sequencing platforms	85
3.3.2.3	Effect of time and diet on microbial gut community profiles	87
3.3.3	Identification of active cellulases.....	95
3.4	Conclusion.....	99
3.5	Acknowledgements	100
3.6	References	100
Chapter 4 Metagenome-based analysis of gut communities in lower and higher termites..... 106		
4.1	Introduction	107
4.2	Methods.....	108
4.2.1	Samples and metagenome sequencing.....	108
4.2.2	Community profiling	108
4.2.3	Metagenome assembly and annotation	108
4.2.4	Gene-centric comparative analysis	109
4.3	Results and Discussion.....	109
4.3.1	Metagenomic sequencing.....	109
4.3.2	Microbial community composition.....	110
4.3.3	Gene-centric analysis of community metabolism.....	115
4.3.4	Functional profiling of the DNA metagenomes.....	115
4.3.5	Plant polysaccharide degradation enzymes.....	116

4.3.6	Other major functions in the termite gut ecosystem	122
4.3.7	Population genome binning	123
4.4	Conclusion.....	124
4.5	Acknowledgements	125
4.6	References	125
Chapter 5	A phylogenomic analysis of the bacterial phylum Fibrobacteres.....	130
5.1	Introduction	132
5.2	Methods	133
5.2.1	Samples and metagenome sequencing.....	133
5.2.2	Sequence assembly and population genome binning.....	133
5.2.3	Taxonomic assignment of population genomes.....	134
5.2.4	Genome annotation and metabolic reconstruction.....	134
5.2.5	Genome and protein family comparative analyses	135
5.3	Results and Discussion.....	135
5.3.1	Recovery of population genomes from environmental metagenomic datasets.....	135
5.3.2	An expanded phylogenetic classification of the phylum Fibrobacteres	136
5.3.3	Inferred metabolism of Fibrobacteres genomes.....	141
5.3.3.1	Polymer hydrolysis.....	141
5.3.3.2	Fermentative metabolism and respiration	149
5.3.3.3	Nitrogen and ammonia metabolism	150
5.3.3.4	Motility and chemotaxis.....	152
5.4	Conclusion.....	152
5.5	Acknowledgements	154
5.6	References	155
Chapter 6	Conclusion and future directions	160
6.1	Conclusion.....	160
6.2	Future directions.....	161
6.3	References	164
Appendix A:	Supplementary figures and tables for Chapter 2.....	166
Appendix B:	Supplementary figures and tables for Chapter 3	188
Appendix C:	Supplementary figures and tables for Chapter 4.....	200
Appendix D:	Supplementary figures and tables for Chapter 5.....	219

List of Figures

Figure 1.1: Phylogenetic tree of termites based on full mitochondrial genome.	5
Figure 1.2: Structural composition of lignocellulosic material.	6
Figure 1.3: A typical termite life-cycle.	7
Figure 1.4: Schematic diagram of termite gut.	9
Figure 1.5: Overview of a typical culture-independent bioinformatic workflow including examples of software used for various steps.	13
Figure 1.6: Example of a metabolic reconstruction, in this case <i>Fibrobacter succinogenes</i>	18
Figure 1.7: Phylum-level diversity of Bacteria identified in the termite gut.	23
Figure 2.1: Heatmap showing microbial taxa (mostly genus and family level) with relative abundance $\geq 0.2\%$ in one or more whole gut samples surveyed in this study.	53
Figure 2.2: Prevalence versus relative abundance graph of bacterial OTUs (97% sequence identity) in the surveyed gut samples.	54
Figure 2.3: Heatmap showing archaeal OTUs (97% seq id) with $\geq 0.1\%$ relative abundance in one or more of the surveyed gut samples.	56
Figure 2.4: UPGMA tree of unweighted (presence/absence only) Soergel pairwise distances between bacterial profiles showing a high consistency with host phylogeny and low consistency with diet.	58
Figure 2.5: Subtrees of host and bacterial community phylogenetic comparisons showing secondary effect of diet on community structure of polyphagous termite genera.	59
Figure 3.1: Photos of a (top) worker <i>M. darwiniensis</i> specimen and (below) dissected <i>M. darwiniensis</i> worker whole gut.	76
Figure 3.2: Photos of 232 x 142 x 64mm containers.	83
Figure 3.3: Photos of worker <i>M. darwiniensis</i> (left) infested with mites and (right) a close up of mite on the head.	84
Figure 3.4: Relative abundances ($\geq 0.2\%$) of <i>M. darwiniensis</i> gut profiles on the <i>E. regnans</i> control diet using different sequencing platforms.	87
Figure 3.5: Phylum-level summary of microbial taxa between wood- and grass-fed <i>Mastotermes</i> . The bar graph represents the averaged of three replicates per treatment.	88
Figure 3.6: Heatmap showing microbial OTUs with $\geq 0.1\%$ relative abundance in the <i>Mastotermes</i> profiled gut samples in feeding trial 2.	91
Figure 3.7: Principle component analysis (PCA) plots of microbial profiles obtained from <i>Mastotermes</i> gut samples biological replicates at 0 hr, Day 4 and Day 7.	92

Figure 3.8: Principle component analysis (PCA) plots of microbial profiles obtained from <i>Mastotermes</i> gut samples average of biological replicates at 0 hr, Day 4 and Day 7.....	95
Figure 3.9: Zymograms of crude extract from <i>M. darwiniensis</i>	96
Figure 3.10: Zymogram (left) and corresponding reference gel (right) of crude extract from <i>M. darwiniensis</i>	97
Figure 3.11: Zymography of endoglucanases produced by <i>M. darwiniensis</i> gut communities.	98
Figure 3.12: Endoglucanase activity (U/mg protein) measured in gut crude extract of termites <i>M. darwiniensis</i> subjected to different diet.	99
Figure 4.1: Prokaryotic community composition at phylum level by metagenome (top) and pyrotag genes (bottom).....	112
Figure 4.2: Comparison of SSU rRNA gene metagenome-based (metagenome) and amplicon (pyrotag) microbial community profiles based on the top 70 OTUs by total relative abundance across all samples.....	113
Figure 4.3: Relative abundance of COG functional categories (% of total coding sequences) across the investigated metagenomes.	118
Figure 4.4: Distribution of genes encoding GH family proteins from the termite gut.	121
Figure 4.5: Distribution of GH families with known hemicellulase activity in this study and other termite gut.	121
Figure 5.1: Phylogenetic analysis of the phylum Fibrobacteres.....	145
Figure 5.2: Glycoside hydrolase families with a significant difference in mean proportions $\geq 1\%$ between Fibrobacteres and other cellulolytic bacteria and a $p \leq 0.05$	149
Figure 5.3: Composite metabolic reconstruction of members of the phylum Fibrobacteres.....	151

List of Tables

Table 1.1: Number of termites in Australia	8
Table 1.2: Examples of some commonly used publicly available assemblers.	14
Table 1.3: A compilation of culture-independent termite guts symbiont research	21
Table 2.1: Summary of the surveyed 66 termite whole gut samples.	49
Table 2.2: Summary of core and accessory bacterial phyla in higher and lower termite gut communities present at >1% relative abundance in at least one sample.	51
Table 3.1: Workers collected at multiple time points from feeding trial 1 and 2.	78
Table 3.2: Description of protein hits as determined by mass spectrometry analysis of trypsin digested gel bands from zymogram gel (Z) and Coomassie-stained gel (C)	97
Table 4.1: Summary statistics of metagenomes	110
Table 5.1: Summary statistics of Fibrobacteres genomes	138
Table 5.2: Inventory of glycoside hydrolases (GHs) identified in the Fibrobacteres genomes, organised by functional category.	139
Table 5.3: Inventory of accessory attachment genes for polymer hydrolysis identified in the Fibrobacteres genomes, organised by Carbohydrate-binding modules (CBMs)	147
Table 5.4: Inventory of accessory attachment genes for polymer hydrolysis identified in the Fibrobacteres genomes, organised by functional category.	148

List of Abbreviations

ARDRA	Amplified Ribosomal DNA Restriction Analysis
BLAST	Basic Local Alignment Search Tool
CAZy	Carbohydrate-Active enZYmes
CBM	Carbohydrate-Binding Modules
CMC	Carboxymethyl Cellulose
COG	Cluster of Orthologous Groups
CDS	Coding Sequences
DGGE	Denaturing Gradient Gel Electrophoresis
DNA	DeoxyriboNucleic Acid
EMP	Embden-Meyerhof-Parnas
ESOM	Emergent Self-Organising Maps
FISH	Fluorescence in situ Hybridisation
GH	Glycoside Hydrolase
HMM	Hidden Markov Model
IMG/ER	Integrated Microbial Genomes Metagenome Expert Review
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC-MS	Liquid Chromatography–Mass Spectrometry
MCP	Methyl Accepting Chemotaxis Protein
ORF	Open Reading Frame
OTU	Operational Taxonomic Unit
PERMANOVA	Permutational Multivariate Analysis of Variance
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PCA	Principal Component Analysis
QIIME	Quantitative Insights Into Microbial Ecology
RFLP	Restriction Fragment Length Polymorphism
rRNA	ribosomal Ribonucleic Acid
SSU	Small SubUnit
TRFLP	Terminal Restriction Fragment Length Polymorphism
TMAH	TetraMethylAmmonium Hydroxide
TCA	TriCarboxylic Acid

Chapter 1 Introduction and Literature Review

1.1 Insect-microbial interactions

Microorganisms evolved and formed spatially organised communities as early as 3.25 billion years ago, well before the appearance of large multicellular organisms (Allwood et al., 2006; Ley et al., 2008). They interact not only with each other, but with the environment as well. Today, diverse microbial communities are widely distributed over the biosphere, from Antarctic sea ice to the human body. Microorganisms play a central role in the regulation of ecosystem processes and have played key roles in driving the evolution of animal and plant life (Ley et al., 2008). The importance of understanding the composition, functional potential and activities of these communities within and across ecosystems is crucial for the survival of eukaryotes as we continue to depend on them. The concept of microorganisms shaping the evolution of their host is long known and in fact, Leigh van Valen, an evolutionary biologist, coined the term Red Queen Hypothesis where he postulated that a constant arms race between prey and host is an important factor in driving evolution (Van Valen, 1973).

1.1.1 Symbiosis

Symbiosis, or “living together” (de Bary, 1879), involves an interaction between two unlike organisms in close association over long periods (Bonfante et al., 2010). It is more diverse in complex systems, where there are networks of interactions of multiple organisms. The traditional concept of symbiosis has focused on the interactions based on nutrition and defence, recognising interactions such as parasitism (one member may be injured from the relationship), commensalism (members are often unaffected) and mutualism (members may benefit) (Bonfante et al., 2010). As microorganisms form relationships with their eukaryotic hosts, they are often referred to as symbionts (Breznak, 2004; Bonfante and Genre, 2010; Bonfante et al., 2010). Microbes that live inside and outside of the host species are referred to as endosymbionts (or inside the cells of the host) and ectosymbionts respectively (Breznak, 2004). These resident microorganisms can influence their hosts at two phases; (1) in “physiological time” that takes into account the composition, density and activities of colonising microorganisms thereby affecting the physiological state of the animal; and (2) in “evolutionary time” that influences the degree of host survival to the infecting microorganisms (Douglas, 2011). The understanding of host-microbe interactions requires the combination of these two perspectives (Douglas, 2011). Symbioses provide unparalleled opportunities for genetic exchange between host and symbiont, often increasing co-dependency, the most extreme version of which are organelles resulting from endosymbiosis (Margulis, 1981).

Margulis emphasised the importance of symbiotic interactions in addition to random mutations and horizontal transfer of metabolic properties between host and symbiont (Bonfante et al., 2010). Over the past decade, symbiosis studies have highlighted the benefits of both or at least one member from interactions (Moran, 2001; Bonfante et al., 2010; Engel et al., 2012; McFall-Ngai et al., 2013). Symbiosis is considered a driving force in shaping the evolution and diversity of eukaryotic organisms, hence leading to biological novelties (Moran, 2006, 2007; Bonfante et al., 2010).

1.1.2 Benefits of resident microbiota on the insect host

A common feature that many hosts such as molluscs and mammals share is that they are colonised by a community of microorganisms that influence their physiological health (McFall-Ngai et al., 2013). Insects as a group have a broader diversity of interactions (e.g. plant-insect, insect-insect, insect-mammal) but contain relatively less gut microbial species (e.g. aphids, honey bee) in contrast to mammals (Douglas, 2011; Engel and Moran, 2013). A variety of insect species is involved in symbiosis (Kaufman et al., 2000; Moran, 2002; Moran, 2007; Brune, 2009). Insects often depend on their mutualistic resident microorganisms for successful growth and reproduction. Microorganisms' metabolic capabilities are valuable assets for insects, mainly in nutrient biosynthesis and biomass degradation (Moran, 2001; Douglas, 2009; Shi et al., 2010; McFall-Ngai et al., 2013). Many insects feed on poor diets that lack essential nutrients such as nitrogen (Mullins and Cochran, 1975; Hongoh and Ishikawa, 1997). The importance of symbionts in nutrient provisioning, that is, production of vitamins, processing food and recycling usable nitrogen in insects has been demonstrated through feeding experiments (Douglas, 2009; Shi et al., 2010). Microbial symbionts in insects are also involved in biomass digestion. Insects and symbiotic microbes are both capable of producing and secreting enzymes responsible for biomass degradation and hydrolysis (Ohkuma, 2003; Tokuda and Watanabe, 2007; Warnecke et al., 2007). Yet, it has been controversial as to whether the insect host or microbiota plays a more significant role in symbiotic digestion (Shi et al., 2010). In recent years, genomic studies have paved the way to explore the critical roles of gut microbiota in the digestion of recalcitrant plant polymers and nutrient provisioning (Santo Domingo, 1998; Cox-Foster et al., 2007; Guan et al., 2007; Aylward et al., 2012; Engel et al., 2012). Studying insect symbioses, from an application point of view, will no doubt enable identification of novel biocatalysts for next-generation biorefinery and development of new tactics for pest management.

1.2 Termites as a model insect for basic and applied research

Termites are important detritivores that contribute significantly to the turnover of lignocellulose in terrestrial ecosystems and to global greenhouse gas production (Bignell et al., 1997). They have been identified as one of the most abundant and ecologically significant eusocial insects (Thorne, 1997; Engel et al., 2009). They are regarded as the most successful wood destroying pests on Earth and negatively impact the timber industry (Verma et al., 2009). An estimated one billion dollars are spent worldwide annually on the control and repair of damage caused by termites (Lax and Osbrink, 2003; Verma et al., 2009). In Australia alone, approximately 130,000 buildings are damaged yearly with costs amounting to 910 million dollars (Hadlington and Staunton, 2008). The enhanced ability of termites to degrade lignocellulose evolved from mutualistic symbioses with their resident gut microbiota (Warnecke et al., 2007; Brune, 2009; Hongoh, 2010). Unlike most other insects, termites have complex and distinctive gut microbial communities responsible for major metabolic processes, including degradation of lignocellulose, homoacetogenesis and nitrogen fixation and recycling (Brune and Ohkuma, 2011). Many termite species can be successfully reared in the laboratory, making them amenable for experimental studies (Salmassi and Leadbetter, 2003).

1.2.1 Evolution and diet

Termites are estimated to have arisen about 170 million years (MYR) (**Figure 1**) from eusocial roaches (Bourguignon et al., 2015). They belong to the order Isoptera which are most closely related to wood-feeding cockroaches (order *Blattaria*) genus *Cryptocercus* (Lo et al., 2000). They are a large and diverse group encompassing an estimated 3,106 known species in 295 genera worldwide (Kambhampati and Eggleton, 2000; Krishna et al., 2013). Termites were previously divided into six extant families, but have been recently reclassified into nine families; the Mastotermitidae, Kalotermitidae, Archotermopsidae, Hodotermitidae, Stolotermitidae, Stylotermitidae, Serritermitidae, Rhinotermitidae, and Termitidae (**Table 1.1**) (Krishna et al., 2013). The first eight families are collectively referred to as lower termites that harbour populations of both cellulolytic flagellated protists and prokaryotes in their guts (Hongoh, 2010). The last family, Termitidae, are also known as higher termites and comprise ~85% of described termite species to date (Kambhampati and Eggleton, 2000; Hongoh, 2010). Higher termites lack flagellated protists which were estimated to be lost from the Termitidae ancestor ~54 MYR (Brandl et al., 2007) in response to dietary diversification (*see below*).

Termites are often perceived as obligate wood feeders but they are more versatile than previously thought. Higher termites have diversified lignocellulose diets ranging from leaves, grass, roots, soils

and dung depending on species (Wood and Johnson, 1986). Some identical higher termite species can feed on different materials are referred to as polyphagous species (i.e. *Amitermes* and *Gnathamitermes*). While those congeneric species that can feed on different materials but each species thrives on a single food type are referred to as monophagous species. Lignocellulose constitutes the bulk of plant biomass, providing rigidity and structure. It is the most abundant organic matter on Earth and a valuable resource for the biofuel industry (Wyman and Yang, 2009). Efficient chemical degradation of lignocellulose is a challenging process due to the complexity of its structure (Rubin, 2008; Sanderson, 2011). Lignocellulose is made up of three carbon-based polymers, cellulose (38-50%), hemicellulose (23-32%) and lignin (15-25%) (Beguin, 1990), with cellulose and lignin identified as the first and second most abundant fixed carbon sources (Vanwonterghem et al., 2014b). Cellulose is comprised of repeating beta (1-4) –linked chains of glucose molecules forming a crystalline structure. Hemicellulose is made up of various hexoses, pentoses and sugar acids that branch around cellulose through hydrogen bonds. Lignin is composed of phenolic compounds that form a protective sheath around the polysaccharides making the biomass resistant to enzymatic digestion (Breznak and Brune, 1994). At present, there is a need for alternative sources of enzymes that can efficiently break down lignocellulose to simple sugars and eventually to biofuel (Sanderson, 2011; Chaturvedi and Verma, 2013).

1.2.2 Australian termites

In Australia, there are approximately 40 termite genera with 268 species that represent five out of the nine families, the Mastotermitidae, Archotermopsidae, Kalotermitidae, Rhinotermitidae and Termitidae (Peters et al., 1996; Hadlington and Staunton, 2008; Constantino, Last updated: September 2012) (**Table 1.1**). They are geographically distributed across the Australian continent, but are typically found in the tropics. The termite nests are often indicative of the termite species which may take form of ground mounds, subterranean or arboreal nests. Interestingly, the habits of some species vary in different climates, for example, *Coptotermes acinaciformis* build mounds in Queensland and not in New South Wales (Hadlington and Staunton, 2008). Of the hundreds of identified Australian species, only a handful is considered pests that infest buildings and forestry. The most common pest species that have been identified are *Mastotermes darwinensis*, *Cryptotermes brevis*, *Porotermes adamsoni*, *Coptotermes acinaciformis*, *Coptotermes frenchi*, *Heterotermes ferox*, *Nasutitermes exitiosus* and *Schedorhinotermes intermedius* (Creffield, 1996; Hadlington and Staunton, 2008). Some subterranean species survive by maintaining nests beneath buildings and pavements in city and suburban areas (Hadlington and Staunton, 2008).

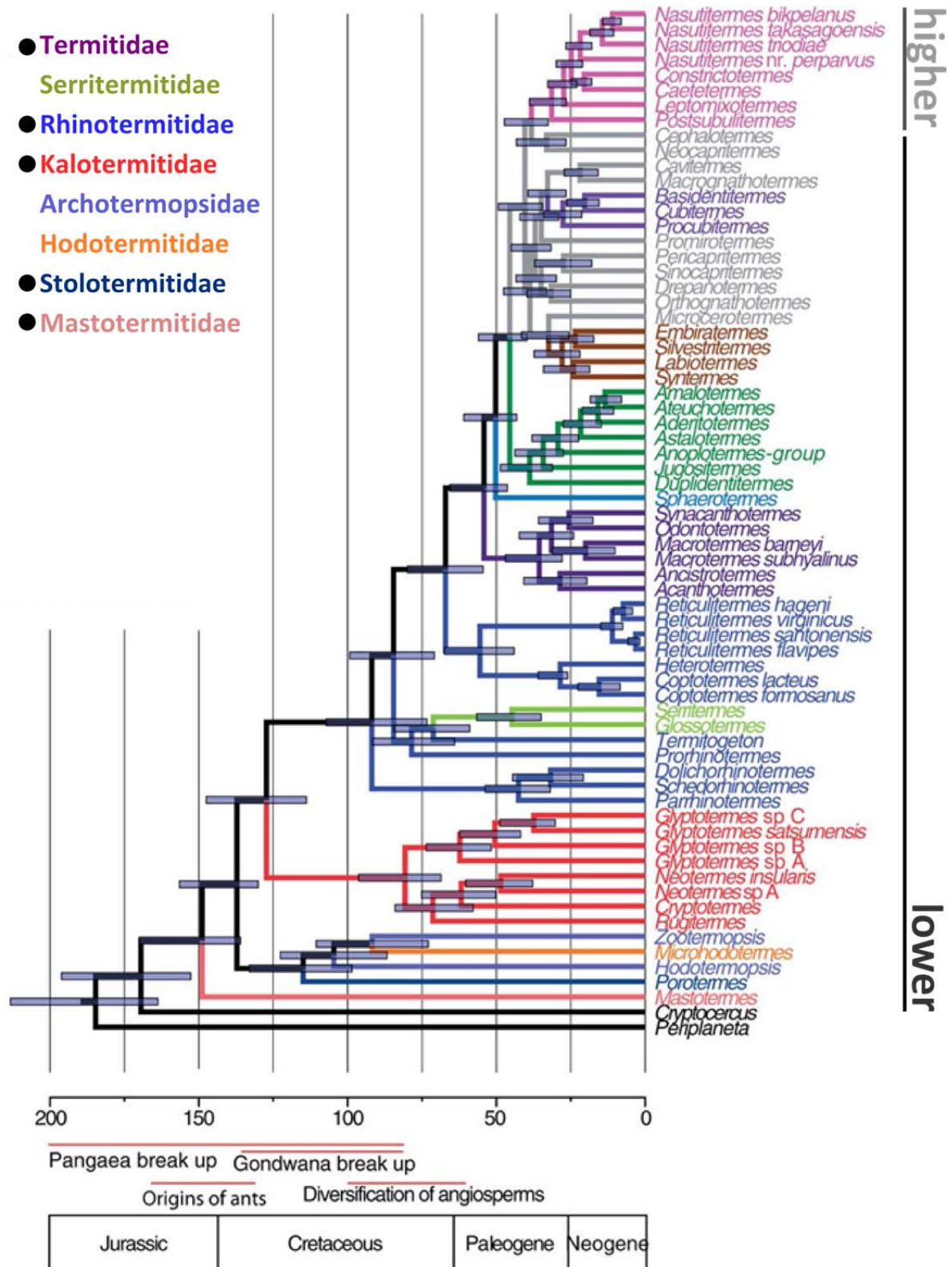


Figure 1.1: Phylogenetic tree of termites based on full mitochondrial genome. The colours of termite families on the left legend correspond to termite genus/species on the tree. Black circles represent termite families included in this thesis. Reproduced from Bourguignon et al. (2015).

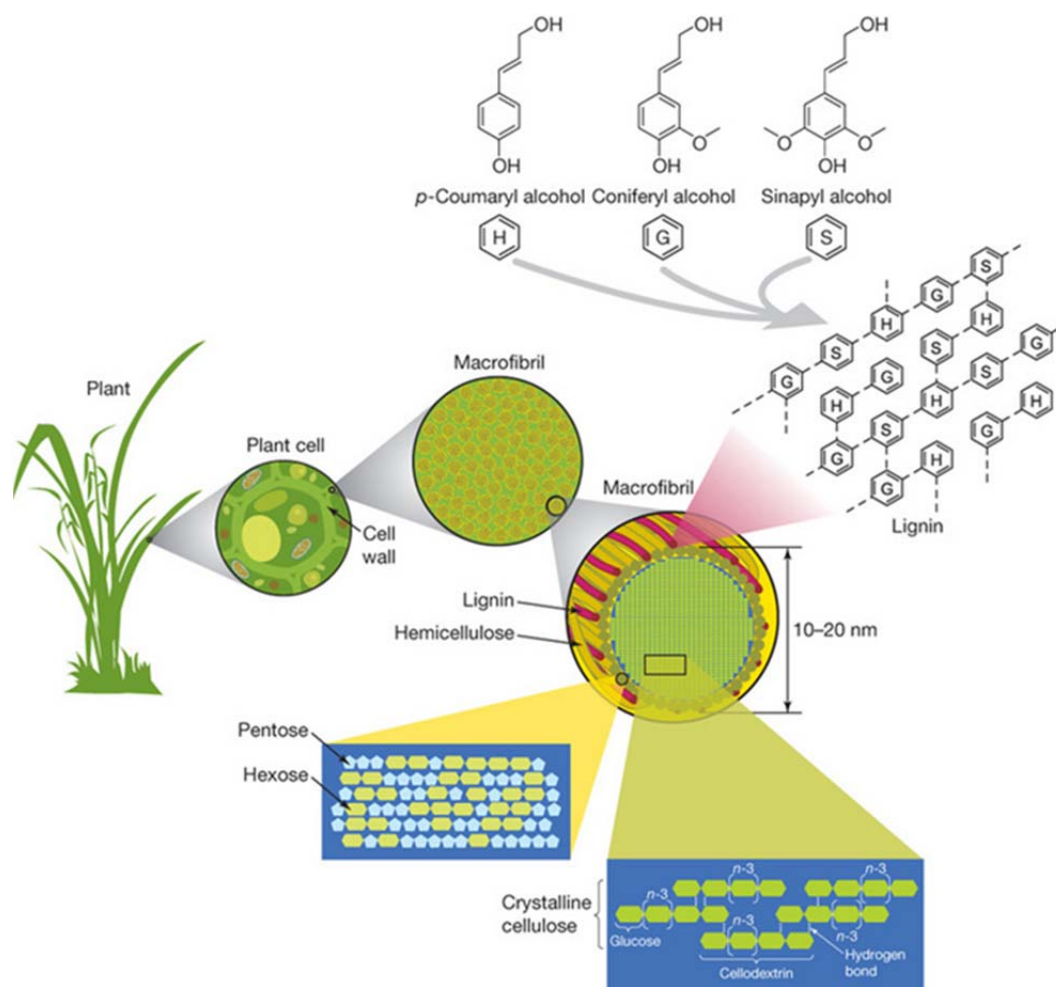


Figure 1.2: Structural composition of lignocellulosic material. Cellulose is protected by tight layers of hemicellulose, pectins chains and lignin. Cellulose makes up 38-50%, hemicellulose 23-32% and lignin 15-25%. Reproduced from Rubin (2008).

1.2.3 Ecology

Like ants, bees, wasps and some aphids, termites are highly social insects that live together in a rigid caste system within the same nest or colony (Davis et al., 2009). There are several castes within the termite colony, most notably workers and soldiers (**Figure 1.3**). Workers have been heavily exploited for studies relating to gut symbionts as they are the most abundant caste and perform the majority of the lignocellulose degradation (Berchtold et al., 1999; Ohkuma et al., 1999; Graber and Breznak, 2004; Hongoh et al., 2005; Warnecke et al., 2007). Workers are also responsible for maintaining major processes in the colony, including building nest and tunnels, hunting for food and water sources, feeding and caring for the other castes (Roisin, 2000; Hadlington and Staunton, 2008).

Termites are broadly classified according to their habitat preference; subterranean, dampwood and drywood (**Table 1.1**) (Hadlington and Staunton, 2008). Subterranean termites are the most abundant

and widely distributed. They often move from parental colony to food sources via construction of underground tunnels. Dampwood termites live in areas where wood is in contact with moist soil such as trees and stumps. Drywood termites inhabit dead timber where atmospheric moisture is often high (Peters et al., 1996; Hadlington and Staunton, 2008).

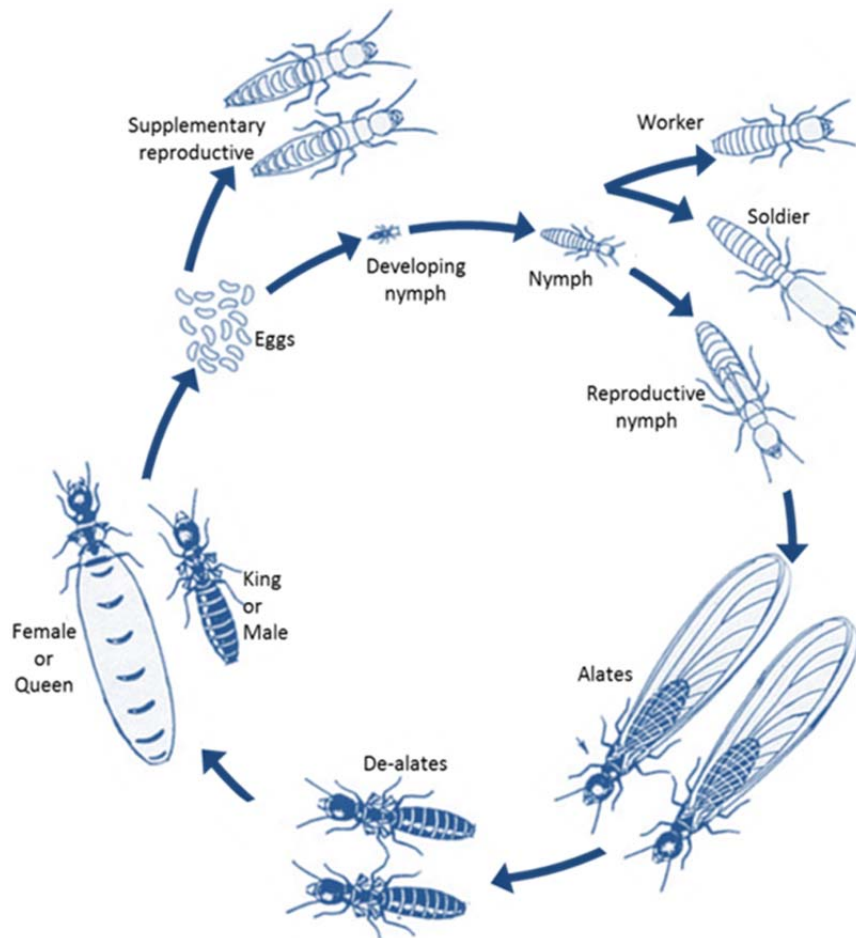


Figure 1.3: A typical termite life-cycle.

Table 1.1: Number of termites in Australia. Adapted by Constantino (Last updated: September 2012)

Family	Sub-Family	Habitat	No. of Genera	No. of Species
<i>Lower termites</i>				
Mastotermitidae	-	Subterranean	1	1
Kalotermitidae	-	Drywood	8	37
Archotermopsidae	-	Dampwood	2	6
Hodotermitidae	-	Subterranean	0	0
Stolotermitidae	-	-	0	0
Stylotermitidae	-	-	0	0
Serritermitidae	-	-	0	0
Rhinotermitidae	-	Subterranean	5	26
<i>Higher termites</i>				
Termitidae	Apicotermitinae	Subterranean	0	0
	Cubitermitinae	-	0	0
	Foraminitermitinae	Subterranean	0	0
	Macrotermitinae	Subterranean	0	0
	Nasutitermitinae	Subterranean	6	44
	Sphaerotermitinae	Subterranean	0	0
	Syntermitinae	Subterranean	0	0
	Termitinae	Subterranean	18	154

1.2.4 Diversity and distribution of microorganisms in the termite gut

Mutualistic relationships between host and gut microbes have enabled animals and herbivorous insects including termites to thrive on recalcitrant plant materials. Many factors such as diet, pH, geographical location, host specificity, and life stage are likely to influence insect gut communities (Robinson et al., 2010; Colman et al., 2012). These factors are not necessarily exclusive of each other. There is evidence that diet and taxonomy of the host can strongly influence insect gut microbial communities (Muegge et al., 2011; Wu et al., 2011; Colman et al., 2012). Like all animals, termite guts are sterile when they hatch and are soon colonised via trophallaxis, which is the transfer of regurgitated or defecated gut contents between colony members (Huang et al., 2008). Termite guts harbour an abundant and diverse symbiotic microbial community comprised of bacteria, archaea and in the case of lower termites, flagellated protists.

Despite its small size, the termite gut has a large surface area in proportion to volume. The termite digestive tract, similar across insects, comprises three main sections; foregut (crop), midgut and hindgut (Engel and Moran, 2013). Ingested woody materials are masticated to smaller fragments by the mandibles before arriving in the crop (Muegge et al., 2011). In 1940, Hungate observed incomplete digestion of sawdust in the crop and midgut of lower termites (Brune, 2014), which are typically smaller than the hindgut. Initial cellulose hydrolysis is initiated by host cellulases secreted by either salivary glands in lower termites or the midgut epithelium in higher termites (Watanabe et

al., 1998; Tokuda et al., 2004). The hindgut is a dilated compartment that houses an estimated 95% of the resident gut microorganisms (**Figure 1.4**) (Berchtold et al., 1999) and where nutrient absorption takes place (Breznak and Brune, 1994).

Termite hindgut microorganisms are responsible for major metabolic activities where they aid in digestive processes; (1) hydrolysis of structural polymers of plant cell walls (cellulose and hemicelluloses), (2) fermentation of the depolymerisation products into acetate (homoacetogenesis) and other short-chain fatty acids which are subsequently reabsorbed by the host, and (3) re-cycling of nitrogenous compounds (Breznak and Brune, 1994; Brune and Friedrich, 2000; Brune, 2014). Termites harbour a distinctive gut microbiota shaped through diet and co-evolution with their hosts (Brune and Ohkuma, 2011).

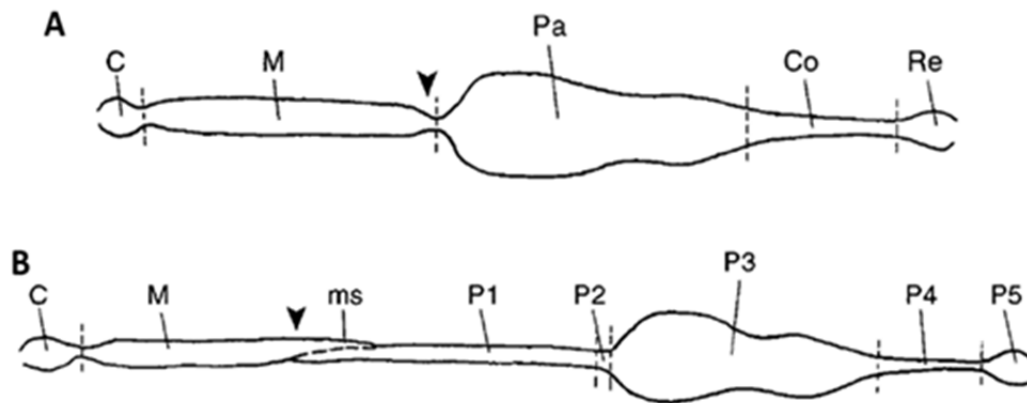


Figure 1.4: Schematic diagram of termite gut. (A) *Reticulitermes flavipes* (Lower termite); C: Crop, M: Midgut, Pa: Paunch, Co: Colon, Re: Rectum (B) *Nasutitermes* species (Higher Termite); C: Crop, M: Midgut, ms: Mixed segment, P1: Anterior to enteric valve, P2: Enteric valve region, P3: Paunch, P5: Rectum. Reproduced from Brune et al. (1995).

1.2.3.1 Symbiotic gut flagellates

Flagellated protists occupy 90% of the hindgut volume of lower termites. It is estimated that 10^3 to 10^5 protists are present in a single gut (Hongoh, 2010). These gut flagellates are unique to termites and are closely related to those in the wood-feeding roach, *Cryptocercus*, sister group of the termite lineage (Ohkuma et al., 2009). They belong to either the phylum Preaxostyla (consisting of one order Oxymonadida) or Parabasalia (consisting of at least four orders Trichonymphida, Spirotrichonymphida, Cristamonadida, Trichomonadida) (Noda et al., 2007; Decker et al., 2009; Brune and Ohkuma, 2011). It is thought that these protist lineages have become specialist wood degraders over the course of 150 MYR of symbiosis with termites and *Cryptocercus* cockroaches (Hongoh, 2010).

1.2.3.2 Symbiotic gut bacteria and archaea

There are approximately 10^6 to 10^8 prokaryotic cells found in a single gut of both lower and higher termites (Schultz and Breznak, 1978; Tholen et al., 1997; Hongoh et al., 2005; Hongoh et al., 2006a). Archaea comprise only a few percent of the total prokaryotes, while the majority are bacteria (Leadbetter and Breznak, 1996; Brauman et al., 2001). The majority of bacterial species within termite guts are thought to be autochthonous symbionts that are vertically transferred (e.g. inherited from other members of the colony) via proctodeal trophallaxis (Hongoh et al., 2005; Berlanga et al., 2009; Hongoh, 2010). It is likely that the majority of bacterial species has coevolved as a community with the host species, with evidence that bacterial phylotypes (community structure) of termite genera have the tendency to cluster together (Hongoh et al., 2005; Brune and Dietrich, 2015). Studies have proposed that bacteria do not strictly co-speciate with termite hosts and may be transferred between termite species belonging to the same genus (Hongoh et al., 2006a; Hongoh et al., 2006b; Hongoh, 2010; Strassert et al., 2010).

1.3 Culture-independent molecular approaches to study microbial ecosystems

Microorganisms are mostly indistinguishable morphologically, making it impossible to classify and reliably identify microbial species based on morphology alone (Amann et al., 1995). Up until the late 20th century, microbiologists depended on microscopy and cultivation as their primary tools in nearly all microbiological studies (Robinson et al., 2010). However, it is estimated that >99% of bacteria observed microscopically are not easily culturable (Amann et al., 1995; Hugenholtz et al., 1998). This phenomenon is known as the “great plate anomaly” (Staley and Konopka, 1985). Culture-independent methods have been developed to investigate microbial communities in diverse environments as an approach to overcome this bottleneck (Staley and Konopka, 1985; Amann et al., 1995; Su et al., 2012). Beginning with the small subunit (SSU) ribosomal RNA (rRNA) gene (16S rRNA and 18S rRNA) in the mid-1980s (Pace, 1997) and progressing to whole genome approaches; metagenomics and other omic techniques (Handelsman, 2004; Cardenas and Tiedje, 2008), culture-independent molecular techniques have revolutionised microbial ecology and evolution. These developments have been driven by rapid improvements in high-throughput sequencing and computing and will be described briefly in turn below.

1.3.1 SSU rRNA-based methods

With the introduction of the polymerase chain reaction (PCR) and sequencing of small subunit (SSU) ribosomal RNA (rRNA) genes (16S rRNA and 18S rRNA) in the mid-1980s, microbial

ecologists have extensively used rRNA sequences to identify bacteria from bulk DNA extracted from environmental samples, thereby bypassing cultivation (Pace et al., 1986; Woese, 1987). The SSU rRNA gene has been considered as the ‘gold standard’ phylogenetic marker for examining the diversity and taxonomic composition of microbial communities (Lane et al., 1985; Ward et al., 1990; Hugenholtz et al., 1998; Case et al., 2007; Tringe and Hugenholtz, 2008), taking advantage of its universal distribution in cellular lifeforms and that it contains both conserved and hypervariable regions suitable for different levels of phylogenetic resolution (Amann et al., 1995; Hugenholtz et al., 1998). PCR provides a rapid means of profiling the constituents of a microbial community via the use of PCR primers that target highly conserved regions of the SSU rRNA gene (Pace, 1997). SSU rRNA genes have been used in a variety of applications to investigate microbiomes including (1) PCR-based fingerprinting methods such as denaturing gradient gel electrophoresis (DGGE) (Muyzer and Smalla, 1998) (2) PCR-clone libraries, and (3) oligonucleotide probe-based hybridisation methods such as fluorescence *in situ* hybridisation (FISH) (Stokes et al., 2001; Su et al., 2012).

More recently, microbial communities have been profiled by multiplexed SSU rRNA gene amplicon sequencing on high throughput next generation (“nextgen”) sequencing platforms (Tringe and Hugenholtz, 2008). The first sequencing technology used for this purpose was Roche 454 pyrosequencing which provided much higher community coverage per sample than the more traditional Sanger sequencing of PCR-clone libraries (Sogin et al., 2006). Within a single sequencing run, hundreds of thousands of SSU rRNA amplicons, referred to as pyrotags, are surveyed from multiple samples in parallel (Ward et al., 2009). Despite this critical breakthrough, pyrosequencing is essentially extinct and has been superseded by other technologies, most notably Illumina sequencing by synthesis technology which produces two orders of magnitude more data for the same cost (Smriga et al., 2010; Caporaso et al., 2011; Loman et al., 2012). There are a number of important methodological caveats with respect to SSU rRNA-based community profiling including primer bias that can affect species evenness (Engelbrektson et al., 2010), sequencing error, particularly for nextgen platforms, that can result in overestimation of species diversity (Kunin et al., 2010), lower phylogenetic resolution of nextgen sequencing reads due to shorter length than Sanger reads (Pontes et al., 2007), and PCR-generated chimeric sequences which inflate diversity estimates and confound phylogenetic tree inference (Pontes et al., 2007). Although SSU rRNA has been pivotal in providing insights into microbial diversity in the environment, revealing much greater species diversity than previously recognised (Simon and Daniel, 2011; Su et al., 2012; Willner and Hugenholtz, 2013), it provides very limited functional information on the profiled microorganisms.

1.3.2 Genome-based methods

Metagenomics is the direct culture-independent genomic analysis of microorganisms in environmental samples. In contrast to SSU rRNA sequencing that lacks functional information, this approach allows a relatively unbiased assessment of the community structure, taxonomy and potential metabolic capabilities (functionality) of a microbial community (Hugenholtz and Tyson, 2008). Since the first metagenomic study of an acid mine drainage biofilm (Tyson et al., 2004) and marine surface water from the Sargasso Sea (Venter et al., 2004), numerous other habitats have been investigated using metagenomics including marine (DeLong et al., 2006; Rusch et al., 2007), terrestrial (Tringe et al., 2005; Blaha et al., 2007) and host-associated ecosystems including insect guts (Warnecke et al., 2007; Aylward et al., 2012; Engel et al., 2012). More than 5000 metagenomes are currently available through the IMG/M database (Markowitz et al., 2014). With the rapid advancement of sequencing capacity and computational power, metagenomics has progressed from identifying over- and under-represented gene families between communities, known as gene-centric analysis (Tringe et al., 2005), to reconstructing metabolic potential of individual population genomes (Kunin et al., 2008; Albertsen et al., 2013). A typical metagenomic bioinformatics workflow includes read quality trimming, assembly, binning, annotation (gene calling) and comparative analysis (Kunin et al., 2008) (**Figure 1.5**) briefly described in turn below.

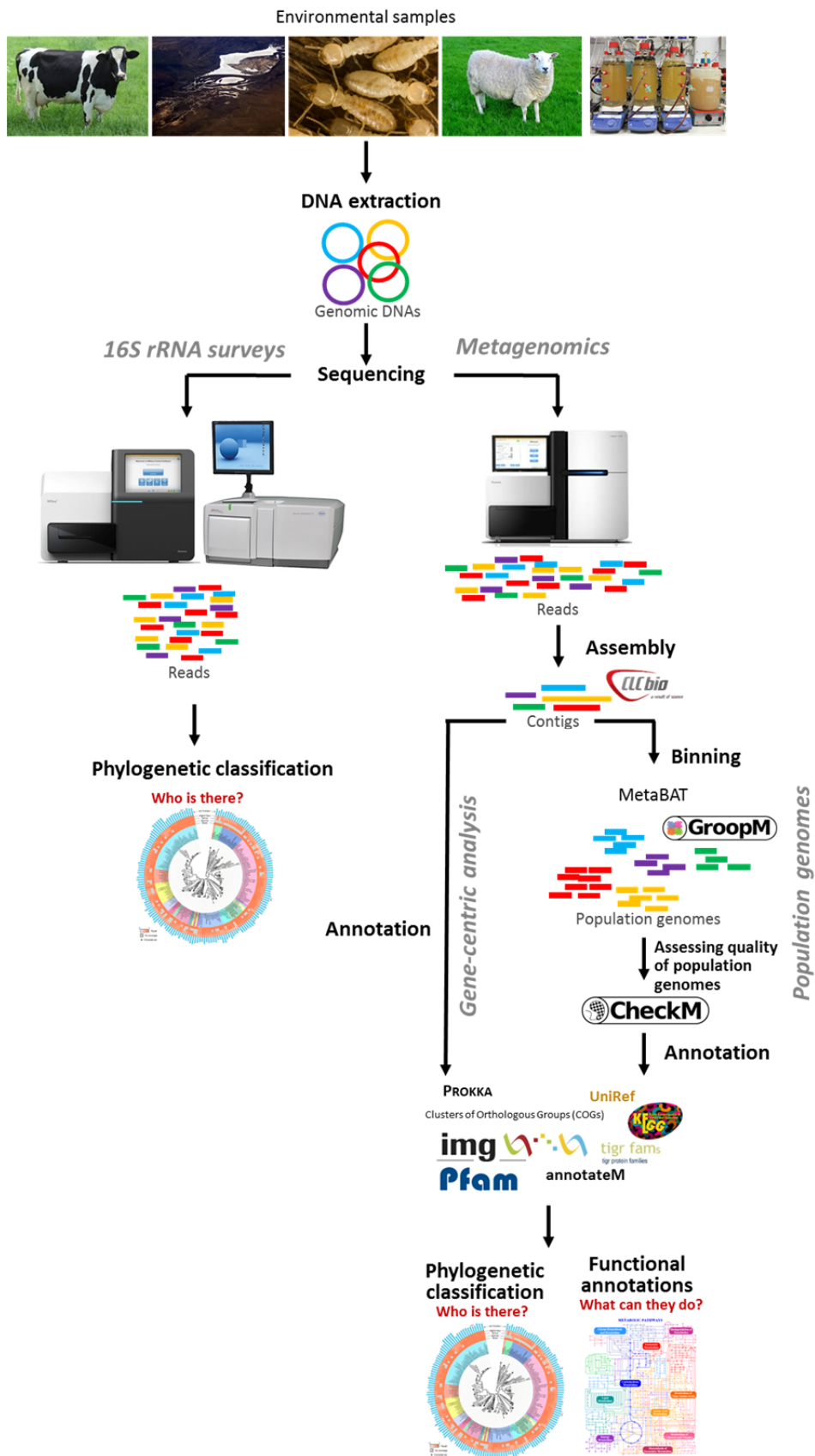


Figure 1.5: Overview of a typical culture-independent bioinformatic workflow including examples of software used for various steps.

1.3.2.1 Reads quality trimming and assembly

One major pitfall of next-generation sequencing is that low-quality sequences typically produce low quality assemblies, including misassembled contigs/scaffolds which can confound subsequent analyses. Therefore, quality trimming sequence reads such as end-trimming sequence adaptors and low quality bases, and removing entire reads that fall below a quality threshold is an important first step in bioinformatics analysis of metagenomic datasets. Read quality is normally assessed based on Phred quality scores which is a statistical estimate of incorrect base calling using programs such as FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads that fall below a common score of ~30, e.g. the probability that 1 in 1000 bases is incorrectly called or 99.9% accuracy, are often excluded (<http://www.phrap.com/phred/>). Sequences that meet the cutoff Phred quality scores are subjected to two-step preliminary trimming via Seqprep (St John, 2013) and Nsoni (<http://vicbioinformatics.com/nesoni.shtml>). Seqprep allows removal of adaptors from high throughput sequencing pipelines while Nsoni performs quality trimming on reads.

Quality trimmed reads are assembled into longer contiguous sequences referred to as contigs using metagenomic assemblers either via reference-based assembly or *de novo* assembly. Contigs can further be connected by paired-end read information into scaffolds (Thomas et al., 2012). Most metagenome assemblers are based on the de Bruijn algorithm that is a representational graph where reads are cut into short sequences (*k*-mers) and capture the overlap ($k-1$ =length of overlap bases) of *k*-mers (**Table 1.2**) (Li et al., 2012; Sharpton, 2014). Some older assemblers are based on overlap-layout-consensus algorithm that calculates pair-wise overlaps between reads and this information is captured on a graph (**Table 1.2**) (Li et al., 2012; Sharpton, 2014). Other short read proprietary assemblers such as CLC workbench (www.clcbio.com) and SeqMan (www.dnastart.com) are commercially available (Miller et al., 2010).

Table 1.2: Examples of some commonly used publicly available assemblers.

Assembler	Assembly algorithm	Reference
Metavelvet	De Bruijn	Namiki et al. (2012)
Ray-Meta	De Bruijn	Boisvert et al. (2012)
Meta-IDBA	De Bruijn	Peng et al. (2011)
Minimus	Overlap-layout-consensus	Sommer et al. (2007)
Newbler	Overlap-layout-consensus	Margulies et al. (2005)
PHRAP	Overlap-layout-consensus	de la Bastide and McCombie (2007)

1.3.2.2 *Binning*

Binning refers to the process of assigning metagenomic sequences into population groups based on either taxonomy-dependent or taxonomy-independent methods (Mande et al., 2012). Taxonomy-dependent based approaches rely on comparisons to known reference sequences and are a form of supervised classification. Comparisons can involve (1) sequence composition (McHardy et al., 2007; Brady and Salzberg, 2009; Rosen et al., 2011) such as GC content, codon usage, oligonucleotide arrangement (Mande et al., 2012), and (2) sequence similarity searches via alignment of reads (e.g. MEGAN, (Huson et al., 2007) or Hidden Markov Models (HMMs) (e.g. CARMA (Krause et al., 2008) and AMPHORA (Wu and Eisen, 2008)). Taxonomy-independent approaches group contigs into bins based on mutual similarities and are a form of unsupervised classification. Similarities can be based on shared tetranucleotide frequencies (sequence composition) and organised into bins via methods such as Emergent self-organising maps (ESOMs) (Ultsch and Mörchen, 2005) and combined with interpolated Markov models (Strous et al., 2012). Recently, another taxonomy-independent approach, differential coverage binning, is starting to gather momentum (Albertsen et al., 2013; Sharon et al., 2013; Imelfort et al., 2014). This approach bin contigs with similar coverage profiles across multiple related metagenomes with the assumption that such contigs belong to the same microbial population. A combination of differential coverage and sequence composition-based binning approaches have been successful in recovering near complete genomes of microbial populations, including rare populations (<1% relative abundance), from a range of habitats. These include activated sludge bioreactors, anaerobic and full-scale wastewater treatment plants, permafrost and faecal communities (Albertsen et al., 2013; Imelfort et al., 2014).

1.3.2.3 *Gene prediction and annotation*

Gene prediction and annotation is a two-step process that involves identifying putative protein, RNA coding (CDSs) and non-coding sequences from contigs or reads (gene prediction) followed by assigning functions to identified genes (annotation). Gene prediction (or gene calling) uses two main approaches, evidence-based or *ab initio* (Kunin et al., 2008). In the evidence-based approach, each metagenomic read is translated into coding sequences by considering all six possible protein coding frames (or known as open reading frames (ORFs)). Subsequently, the resulting peptides are compared to reference databases based on similarity searches using BLASTP or faster algorithms such as USEARCH (Edgar, 2010), RAPsearch (Zhao et al., 2012), or LAST (Kielbasa et al., 2011). This gene prediction method is most often used together with functional annotation (*see below*).

The main caveat of this approach is that novel CDSs may have no statistically significant similarity to database sequences and will consequently be undetected by this approach (Sharpton, 2014).

The *ab initio* method detects CDSs using gene prediction models trained on properties of microbial genes (e.g. GC content, codon usage, gene length, identification of promotor motifs). Several readily available tools such as Prodigal (Hyatt et al., 2010), MetaGene (Noguchi et al., 2006) and GeneMark (Besemer and Borodovsky, 2005) exploit this approach. Even though these tools are based on the same principles, their performance differs as a function of read properties (e.g. length and sequencing error rate) and therefore, research analysts should be mindful in choosing a suitable algorithm for their datasets (Trimble et al., 2012). MetaGene and GeneMark are generally applied to assembled contigs while Prodigal can be used on both metagenomic reads and contigs. In contrast to evidence-based approaches, the *ab initio* gene prediction method allows detection of novel sequences not represented in public databases, which ideally require validation (e.g. expressed mRNA/protein) to prevent spurious calls (Sharpton, 2014). Note that if the reference gene is spurious, similarity-based searches can result in additional spurious calls creating phantom gene families and adding to the phenomenon of ‘annotation rot’.

Once CDSs has been predicted, functional annotation follows. This is achieved by comparing the translated CDSs to either a database of protein sequences or a database of probabilistic gene models, the most widely used being Hidden Markov Models (HMM) (Rabiner and Juang, 1986) of protein families, wherein a protein family consist of a group of phylogenetically related protein sequences (Finn et al., 2014; Sharpton, 2014). Predicted CDSs are compared to all proteins or models, and are then classified into either a single family, a series of families, or no family (in the case of novel, highly diverged or potentially spurious protein sequences). Sequence databases like the SEED annotation system (Overbeek et al., 2005), KEGG ortholog groups (Kanehisa et al., 2011), MetaCyc (Caspi et al., 2010) and software tools such as Prokka (Seemann, 2014) are frequently used for functional annotation (Karp and Caspi, 2011). These programs are often used in parallel to ensure all aspects of metabolic pathways are covered. For example, contigs can be annotated by Prokka and the coding sequences mapped to KEGG pathways. MetaCyc can be used to crosscheck the presence of pathways that may be absent in KEGG. Searching through databases using HMMs can be less accurate due to use of highly conserved sequence motifs leading to classification of distantly related homologs. Pfam (Finn et al., 2014) is an example of a HMM database that uses HMM probabilistic models to search and annotate protein domains. A number of user-friendly web-based programs exist which perform gene prediction, database searches, protein family classifications and annotations on submitted metagenomes in the context of publicly available datasets. These include MG-RAST (Meyer et al., 2008) and IMG/M (Dai et al., 2012).

1.3.2.4 *Metabolic reconstruction and comparative analysis*

The functional interpretation of hundreds to thousands of genes is key to incorporating metabolic potentials of a particular organism and/or microbial community within an ecosystem (Abubucker et al., 2012). Metabolic reconstruction is the process of integrating annotated genes and pathways identified in a genome into a cellular framework; an example is shown in **Figure 1.6**. Surprisingly, no dedicated software exists to perform this process and currently reconstructions are performed manually in reference to the primary literature. Limitations in metabolic reconstruction include (1) ambiguous or incorrect gene annotations including the large fraction of ‘hypothetical’ and ‘putative’ genes found in a typical microbial genome, (2) multiple functions of a given enzyme (e.g. succinate dehydrogenase; EC 1.3.5.1), and (3) missing genes or incomplete pathways or functional modules leading to uncertainty whether the organism is capable of a given function. This can be a particular issue with incomplete genomes derived from metagenomes. Hence, it is important that multiple lines of evidence are pursued to confirm or refute the presence of metabolic functions, such as gene neighbourhood searches to identify potentially overlooked genes, and phylogenetic trees to confirm orthology and therefore function.

Comparative genomics involves the comparison of the metabolic reconstructions of phylogenetically or ecologically related organisms to identify commonalities and differences which can provide insights on the ecology and evolution of a set of organisms (Hardison, 2003). Note that the process of annotation and metabolic reconstruction already implicitly uses comparative genomics. A wide range of tools exist to facilitate genome comparisons and visualisation. Proteinortho (McDonald et al., 2008) and OrthoMCL-DB (Cantarel et al., 2009) are tools that use reciprocal best alignment matches to identify orthologs and paralogs from multiple species. Visualisation tools such as STAMP provides user-friendly exploration of statistical results with easy output of plots, measurements of effect sizes and confidence intervals (Parks et al., 2014). VISTA is a web-accessible software program offering multiple tools for comparative analysis of genomes including rVista and GenomeVista that use global alignment strategies for annotation and phylo-Vista for analysis and visualisation (Frazer et al., 2004). The aforementioned MG-RAST and IMG/M also support comparative analysis of multiple genomes in the context of extensive reference genome and metagenome databases.

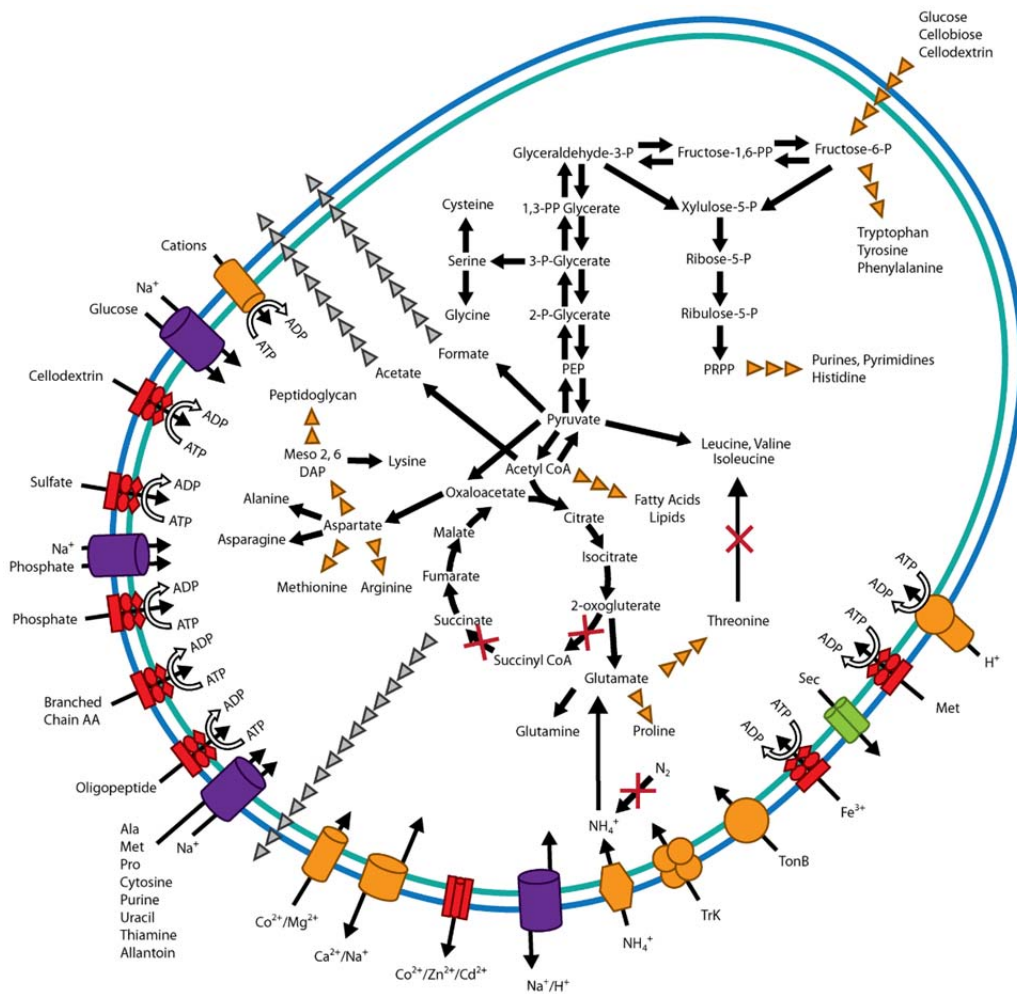


Figure 1.6: Example of a metabolic reconstruction, in this case *Fibrobacter succinogenes*. Reproduced from Suen et al., 2011.

1.3.2.5 Metatranscriptomics and metaproteomics

Metagenomics has provided abundant information on the metabolic and functional potential of uncultured microorganisms, but as a DNA-based approach, it cannot differentiate between expressed and non-expressed genes and is unable to capture the actual metabolic activity of a microbial community (Simon and Daniel, 2011; Su et al., 2012). Metatranscriptomics is the profiling of expressed genes and has been used for identification of RNA-based functional regulation and expressed biological profiles in complex ecosystems (Simon and Daniel, 2011). Metatranscriptomics involves three essential steps: (1) extraction of total RNA, (2) removal of rRNAs to enrich mRNAs, and (3) cDNA synthesis. Extra precaution should be taken when handling RNAs as they are less stable than DNAs (e.g. some transcripts have a life time of less than one minute) (Kan et al., 2005). Enrichment of mRNA is a crucial step since rRNAs make up 95-99% of bulk RNA in microbial cells. Several strategies exist for enriching mRNAs, the two most commonly used are (1) hybridisation of rRNA to sequence-based capture probes, which are subsequently removed from the RNA pool, and (2) targeting the polyA tails of mRNAs, which are

largely absent on rRNAs, using an oligo(dT) primer (Kan et al., 2005). However, mRNA polyA tails are more common in eukaryotes than prokaryotes, so this latter method will miss the majority of prokaryotic mRNAs (Dreyfus and Régnier, 2002).

Similarly, metaproteomics provides insights into functional gene expression but via characterisation of proteins expressed at a given time within an ecosystem (Su et al., 2012). A standard experimental procedure of metaproteomics include: (1) samples preparation such as protein extraction, purification and concentration; (2) proteins denaturation and reduction; (3) protein or peptide separation and quantification, enzymatic digestion and mass spectrometry sequencing; (4) protein identification (Su et al., 2012). Emergence of metatranscriptomics and metaproteomics in recent years has enabled the study of gene expression and protein production of microbial communities to complement metagenomics DNA-based analyses. These approaches provide insights into the functional dynamics of microbial communities, but a number of caveats should be kept in mind. In metatranscriptomic studies, potential limitations include low mRNA recovery due to the need to separate mRNA from the more abundant structural RNA molecules (rRNA, tRNA), change in mRNA profiles due to processing (e.g. storing samples at room temperature), the variable half-lives of different mRNAs, and issues with extrapolating mRNA abundance to protein abundance and activity (Su et al., 2012). Some of these limitations can be overcome through improvements of metatranscriptomic protocols and appropriate processing and storage of samples (Vanwonderghem et al., 2014a). In metaproteomics, challenges include effect of sample collection and storage conditions on quality and quantity of proteins, difficulty in extracting a suitable protein fraction due to the complexity of microbial communities, identification of trypsin-digested peptides, and the extrapolation of protein abundance to activity (Kan et al., 2005; Su et al., 2012). Even so, the combination of “omic” approaches (metagenomics, metatranscriptomics, metaproteomics, etc) offers significant promise to understand the function and dynamics of microbial populations in their natural settings.

1.4 Application of culture-independent molecular methods to termite gut communities

Traditionally, imaging and culture-based methods have been applied to termite gut microbial symbionts (Eutick et al., 1978; O'Brien and Slaytor, 1982). Since most microorganisms cannot be readily cultured (Pace, 1997), including ~90% of all prokaryotes in the termite hindgut, culture-independent techniques are necessary to understand termite gut microbial communities (Brune and Friedrich, 2000). SSU rRNA-based approaches were first applied to study the termite gut in the mid-1990s (Berchtold, 1994; Ohkuma and Kudo, 1996). These included DGGE, terminal restriction

fragment length polymorphism (T-RFLP), and fluorescence *in situ* hybridisation (FISH) (Shi et al., 2010). In recent years, omic approaches have become a preferred option over SSU rRNA-based as they provide functional insights into uncultured gut symbiont populations and communities. Since the first application of metagenomics (Warnecke et al., 2007) and metatranscriptomics (Todaka et al., 2007) to termite guts, meta-omic approaches have increasingly been used to study the relationship of the termite host and its gut microbiome (**Table 1.3**). Culture-independent approaches have revealed a great deal of uncharacterised microbial diversity in this ecosystem and have contributed to our understanding of the composition and function of gut microbiota, which will be discussed in more detail below.

Table 1.3: A compilation of culture-independent termite guts symbiont research. Adapted from Scharf (2015b).

Approach taken	Method ^a	Termite genus	Reference
SSU-based			
<i>16S rRNA genes</i> <i>Bacteria</i>	DGGE	<i>Reticulitermes</i> and <i>Nasutitermes</i>	Bauer et al. (2000)
	T-RFLP	<i>Cubitermes</i>	Schmitt-Wagner et al. (2003)
	T-RFLP	<i>Macrotermes</i>	Hongoh et al. (2006b)
	FISH	<i>Mastotermes</i>	Berchtold et al. (1999)
	FISH	<i>Cubitermes</i> .	Schmitt-Wagner et al. (2003)
	FISH	<i>Microcerotermes</i> and <i>Nasutitermes</i>	Hongoh et al. (2006a)
	Sanger sequencing	<i>Coptotermes</i>	Husseneder et al. (2010)
	Sanger sequencing	Multi-species: <i>Reticulitermes</i> , <i>Hodotermopsis</i> , <i>Zootermopsis</i> , <i>Mastotermes</i> , <i>Kalotermes</i> , <i>Neotermes</i> , <i>Cryptotermes</i>	Stingl et al. (2005)
	Sanger sequencing	Multi-species: <i>Reticulitermes</i> , <i>Coptotermes</i> , <i>Zootermopsis</i>	Lilburn et al. (1999)
	Sanger sequencing + ARDRA analysis	<i>Reticulitermes</i>	Fisher et al. (2007)
	454 pyrosequencing		Boucias et al. (2013)
	Sanger sequencing + T-RFLP analysis	<i>Reticulitermes</i>	Yang et al. (2005)
	Sanger sequencing	<i>Reticulitermes</i>	Hongoh et al. (2003)
	Sanger sequencing	<i>Zootermopsis</i>	Rosengaus et al. (2011)
	Sanger sequencing	<i>Cornitermes</i>	Grieco et al. (2013)
	Sanger sequencing + DGGE	<i>Cubitermes</i>	Fall et al. (2007)
	Sanger sequencing	Multi-species: <i>Odontotermes</i> and <i>Microtermes</i>	Makonde et al. (2013)
	454 pyrosequencing	Multi-species:	Otani et al. (2014)
	454 pyrosequencing	<i>Nasutitermes</i>	Engelbrekton et al. (2010)
	454 pyrosequencing		Köhler et al. (2012)
Sanger sequencing + T-RFLP analysis	<i>Nasutitermes</i>	Husseneder et al. (2010)	
454 pyrosequencing	<i>Trinervitermes</i>	Sanyika et al. (2012)	
454 pyrosequencing	Multi-species: <i>Reticulitermes</i> , <i>Zootermopsis</i> , <i>Cryptocercus</i>	Tai and Keeling (2013)	
<i>18S rRNA genes</i> <i>Protist</i>			
<i>18S and 16S rRNA</i> <i>genes</i> <i>Protist and bacteria</i>	Sanger sequencing	<i>Zootermopsis</i>	Tai et al. (2013)
	454 pyrosequencing	Multi-species: 24 lower termites and three <i>Cryptocercus</i> cockroaches	Tai et al. (2015)
Omics-based			
<i>Genome, fungal</i> <i>symbiont genome, and</i> <i>gut microbial</i> <i>metagenome</i>	Illumina sequencing	<i>Macrotermes</i> (and <i>Termitomyces</i> sp. symbiont)	Poulsen et al. (2014)
<i>Transcriptome</i>	Sanger sequencing	<i>Coptotermes</i>	Hussain et al. (2013)
	Representational difference analysis, Sanger sequencing	<i>Cryptotermes</i>	Weil et al. (2007)
	Sanger sequencing	<i>Reticulitermes</i>	Gao et al. (2012)
	Subtractive cDNA library, Sanger sequencing	<i>Macrotermes</i>	Johjima et al. (2006)
	454 pyrosequencing	<i>Termitomyces</i>	Yang et al. (2012)
<i>Metagenome</i>	Illumina de novo genome sequencing	<i>Coptotermes</i>	Do et al. (2014)
	Functional screening + Sanger sequencing	<i>Coptotermes</i>	Mattéotti et al. (2012) Mattéotti et al. (2011)
	Functional screening (beta glucosidase) + Sanger sequencing	<i>Globitermes</i>	Wang et al. (2012)
	454 pyrosequencing (bacterial fosmid libraries grown under	<i>Macrotermes</i>	Liu et al. (2011)

	selective conditions) Functional screening (cellulase and xylanase) + Sanger sequencing	<i>Microcerotermes</i>	Nimchua et al. (2012)
	454 pyrosequencing	<i>Pseudacanthotermes</i>	Bastien et al. (2013)
	Functional screening (esterase)+ Sanger sequencing	<i>Trinervitermes</i>	Rashamuse et al. (2012)
	Functional screening (feruloyl "FAE" esterase) + Sanger sequencing	<i>Trinervitermes</i>	Rashamuse et al. (2014)
<i>Metatranscriptome</i>	Sanger sequencing (normalised polyphenic library)	<i>Coptotermes</i>	Zhang et al. (2012)
	454 pyrosequencing		Xie et al. (2012)
	Sanger sequencing	<i>Hodotermopsis</i>	Yuki et al. (2008)
<i>Metatranscriptome</i>	Sanger sequencing	Multi-species: <i>Reticulitermes, Hodotermopsis,</i> <i>Neotermes, Mastotermes,</i> <i>Cryptocercus</i> cockroach	Todaya et al. (2010)
	Sanger sequencing (random clones)	<i>Reticulitermes</i>	Wu-Scharf et al. (2003)
	Filter arrays, Sanger sequencing	<i>Reticulitermes</i>	Scharf et al. (2003)
	Filter arrays, Sanger sequencing	<i>Reticulitermes</i>	Scharf et al. (2005)
	Sanger sequencing	<i>Reticulitermes</i>	Tartar et al. (2009)
<i>Metatranscriptome</i>	Sanger sequencing (random clones)	<i>Reticulitermes</i>	Steller et al. (2010)
	Microarray	<i>Reticulitermes</i>	Sen et al. (2013)
	Microarray	<i>Reticulitermes</i>	Raychoudhury et al. (2013)
	Illumina sequencing	<i>Zootermopsis</i>	Rosenthal et al. (2011)
<i>Metabolome</i>	Isotope-ratio mass spectrometry (IR-MS)	<i>Hodotermopsis</i>	Tokuda et al. (2014)
	TMAH thermochemical lysis coupled with GC-MS	<i>Zootermopsis</i>	Geib et al. (2008)
	TMAH thermochemical lysis coupled with CP-MAS-NMR spectroscopy, and Py-GC/MS	<i>Coptotermes</i>	Ke et al. (2011)
	TMAH thermochemical lysis coupled with GC-MS	<i>Coptotermes</i>	Ke et al. (2013)
<i>Proteome</i>	LC-MS/MS (ion trap) and 2-D PAGE	<i>Reticulitermes</i>	Bauwens et al. (2013)
	LC-MS	<i>Nasutitermes</i>	Burnum et al. (2011)
<i>Metatranscriptome and proteome</i>	454 pyrosequencing + LC- MS proteomics	<i>Reticulitermes</i>	Sethi et al. (2012)
Combination of SSU- and genome-based			
<i>Metagenome and 16S survey</i>	454 pyrosequencing	<i>Odontotermes</i>	Liu et al. (2013)
<i>Metatranscriptome, metagenome, and 16S pyrosequencing</i>	454 pyrosequencing	Multi-species: <i>Amitermes</i> and <i>Nasutitermes</i>	He et al. (2013)

^a DGGE, Denaturing gradient gel electrophoresis; T-RFLP, Terminal restriction fragment length polymorphism; FISH, Fluorescence *in situ* hybridisation; ARDRA, Amplified ribosomal DNA restriction analysis ; LC-MS, Liquid chromatography–mass spectrometry; PAGE, Polyacrylamide gel electrophoresis; TMAH, Tetramethylammonium hydroxide; Py-GC/MS, Pyrolysis gas chromatography mass spectrometry.

1.4.1 Discovery of novel bacterial diversity in the termite gut

Characterisation of termite hindgut microbiota using culture-independent methods (**Table 1.3**) has led to the discovery of more than 1500 species of bacteria grouped into 24 phylum-level clusters (**Figure 1.7**) (Hongoh, 2010). Of these clusters, three anaerobic bacterial groups have been reported to dominate the guts of all termites examined to date; (1) the genus *Treponema* belonging to the Spirochaetes, (2) the order *Bacteroidales* in the Bacteroidetes and (3) the class *Clostridia* in the Firmicutes (Eutick et al., 1976; Hongoh, 2010; Brune, 2014). Also, two novel bacterial phyla were first identified in the termite gut (TG), candidate phylum TG1 (later reclassified as the Elusimicrobia) and candidate phylum TG3, which are abundant in lower and higher termites respectively (Hugenholtz et al., 1998; Ohkuma, 2003; Hongoh et al., 2006; Herlemann et al., 2007; Ohkuma et al., 2007).

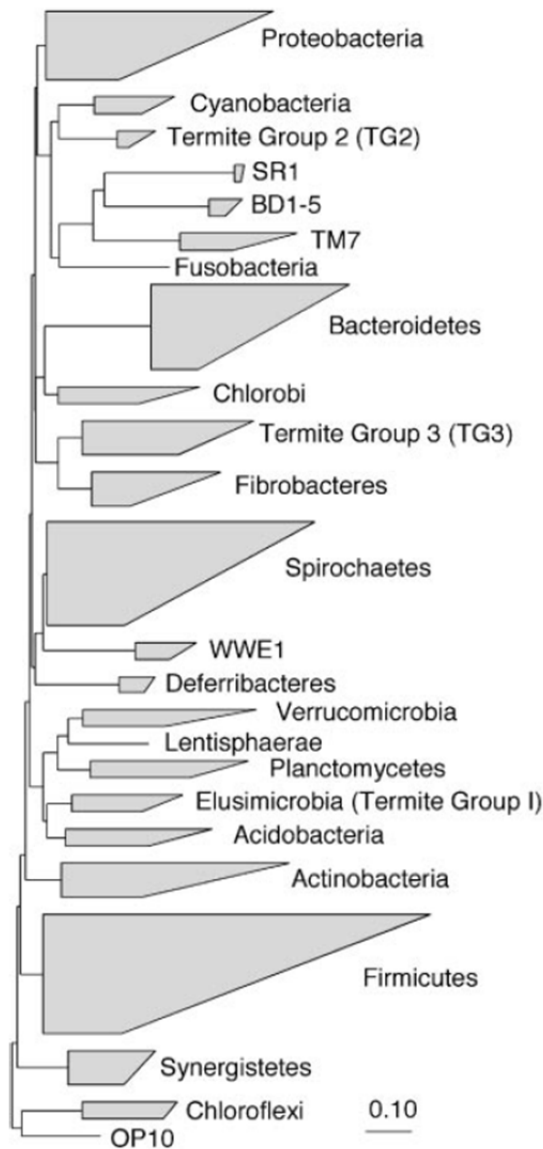


Figure 1.7: Phylum-level diversity of Bacteria identified in the termite gut. Reproduced from Hongoh (2010).

1.4.2 Metagenomic analyses of termite gut microbiomes

‘Omic’ approaches have been applied to termite guts to explore the structure and function of diverse microbial populations involved in the degradation of lignocellulosic biomass (**Table 1.3**) with the greatest emphasis on protist symbionts of lower termites (Scharf, 2015a). The first metagenomic study of termite gut communities confirmed the essential roles of gut symbionts in cellulose and hemicellulose hydrolysis, nitrogen fixation and recycling, and homoacetogenesis (Warnecke et al., 2007). Warnecke et al. (2007) revealed over 700 glycoside hydrolases (GHs) in the hindgut of a higher termite (*Nasutitermes*), comprising 45 carbohydrate active enzyme (CAZy) families, many with recognised activity in degradation of plant cell walls (cellulose and hemicellulose) and a diverse group of gut bacteria representing 12 phyla and 216 phylotypes.

Subsequent metagenomic studies on termite gut symbionts have mainly focused on identification of lignocellulose-degrading enzymes from microbiota genomes. For example, Liu et al. (2013) and He et al. (2013) reported high numbers of bacterial cellulase and hemicellulase genes from the gut metagenomes of higher termite fungus-growing *Odontotermes* (205), wood-feeding *Nasutitermes* (469) and dung-feeding *Amitermes* (574). While these studies and other SSU-based studies have provided evidence for fibre digestion roles of gut bacterial communities in higher termites, fewer efforts have been invested into understanding the contribution of gut bacteria to this process in lower termites. This is because bacteria have been largely overlooked in lower termites in favour of flagellated protists which have been long known as key players in lignocellulose degradation (Brune, 2014). Metagenomic approaches have recently been applied to exploring the enzymatic involvement of bacterial microbiota in lower termites (**Table 1.3**). For example, Do et al. (2014) identified 587 ORFs encoding cellulases, hemicellulases and pectanases from free-living bacterial gut symbionts of *Coptotermes gestroi*, suggesting that bacteria may play a larger role in lignocellulose degradation in lower termites than previously thought. Lignin-degrading genes however remain poorly characterised in both higher and lower termite gut symbiont metagenomic studies because they are either unknown, or belong to generic gene families such as peroxidases, which play multiple roles (Pace, 1997). In addition to targeted searches, gene-centric analysis allows detection of over- and under-represented gene families between termite guts or relative to other ecosystems pointing to potentially important ecosystem-specific functionality. For example, cohesin and dockerin genes, key components of the cellulosome (Todaka et al., 2007), are heavily under-represented in most termite guts (Warnecke et al., 2007). However, He et al. (2013) found a higher abundance of these gene families in the metagenome of dung-feeding *Amitermes* relative to wood-feeding *Nasutitermes*. Together with SSU-based community profiling, they inferred that

cellulosome-containing *Clostridia* populations were laterally acquired by *Amitermes* from herbivore dung.

1.4.3 Expressed genes in termite guts

Metatranscriptomics and metaproteomics have complemented metagenomic studies by confirming expression of genes inferred to be important in termite gut biology (Tartar et al., 2009; He et al., 2013; Raychoudhury et al., 2013) and by providing information on the response of gut communities to different feeding regimens (or alternative biofuel biomass in feeding experiments) (Scharf, 2015b). One of the first metatranscriptomic studies of a termite gut (*Reticulitermes flavipes*) revealed 6555 putative transcript sequences, of which 3044 were inferred to be host-derived and 3511 from protist symbionts (Tartar et al., 2009). Of these, 171 were putative cellulase and hemicellulase genes, 37% from the host and 57% from protists, suggesting that both are involved in polymer digestion through production of cellulases and hemicellulases respectively (Tartar et al., 2009). Homologues of lignase, antioxidant and detoxification enzymes were also detected in the host library, which reported possible involvement in lignin degradation by a termite host. The same method was used by the same group to investigate the differences in gene expression between wood- and paper-fed *Reticulitermes* resulting in identification of a set of host and protist symbiont transcripts with differential abundance between the two diet treatments (Raychoudhury et al., 2013). The majority of transcripts responsive to paper and wood were derived from protists and host respectively, and encoded putative digestive and wood-related detoxification enzymes. In both studies, bacterial transcripts represented only a minority of the total expressed genes, but this may have been due to the use of polyA tailing which would exclude most bacterial transcripts (Tartar et al., 2009; Raychoudhury et al., 2013), leaving open the question of bacterial involvement in lignocellulose degradation in this lower termite. Using rRNA depletion, He et al. (2013) identified bacterial glycoside hydrolases (GHs) and cellulose-binding modules (CBMs) expressed in response to cellulose and hemicellulose digestion from the gut metatranscriptomes of dung-feeding *Amitermes* and wood-feeding *Nasutitermes*, both higher termite genera lacking protists. The transcript profiles were consistent with dietary differences, with higher levels of cellulase and nitrogen fixing gene expression in wood-feeders, and grass side chain-cleaving hemicellulase expression and fixed-nitrogen utilisation in dung-feeders. Importantly, some genes with low relative abundance in corresponding metagenomes had high expression levels, such as GH11s (xylanases), highlighting the potential limitation of gene-centric analysis in metagenome-only studies.

Metaproteomics have also been used to explore system dynamics of termite gut symbionts and in some cases to provide validation of metatranscriptome data (Scharf, 2015a). In the metagenomics

study of a *Nasutitermes* species by Warnecke et al. (2007), a metaproteome of the hindgut luminal fluid was examined to identify which genes were expressed, confirming the inferred importance of cellulases and hemicellulases in this ecosystem. For example, the metaproteome analysis revealed that only one of the 46 endoxylanase (GH10) genes identified was highly expressed at the time of sampling. Metaproteomics has also provided new insights into the long-standing question of whether lignin degradation is occurring in the termite gut. Over 9500 termite host and protist genes were differentially expressed in *Reticulitermes* fed with varying degrees of lignin-complex diets, of which 96 could be connected to lignin degradation (Sethi et al., 2012). Metatranscriptomics and metaproteomics have not been applied extensively to termite gut ecosystems to date (**Table 1.3**), and hold great potential for enhancing our understanding of termite gut microbiology (Shi et al., 2010).

1.5 Summary of results chapters

Since the majority of termite gut microbiota is comprised of unculturable microbes, detailed analysis has been hampered until the introduction of molecular techniques in the 1990s. In recent years, there has been growing interest in understanding termite gut symbionts with newly developed high-throughput technologies. This is in conjunction with the applied goals of advancement in biofuel production and pest management (Hongoh, 2010; Shi et al., 2010). Even after more than 20 years of research, we have only scratched the surface of unveiling the diversity and symbiotic mechanisms of these tiny bioreactors. There are yet insufficient data to answer the broader ecological questions that include the effect of diet versus co-evolution, the effect of changing diet on microbial community structure and function, the function of specific populations, and relative function of prokaryotic and eukaryotic symbionts to hydrolysis in lower termites. Conspicuously, little molecular data exist for the majority of Australian termite species.

This thesis aims to address these questions through SSU profiling, feeding trials, metagenomics, metaproteomics and enzyme characterisation analyses of termite gut microbiota, with a special emphasis on Australian termites. Chapter 2 describes a large molecular survey and comparative analysis of Australian and North American termite microbiota via SSU rRNA gene sequencing. Chapter 3 examines the gut microbial community composition with dietary fluctuations and time through a series of feeding trials on the indigenous Australian termite, *Mastotermes darwiniensis*. Chapter 4 investigates the microbial genomic diversity and functionality by performing shotgun sequencing of the gut of four wood-feeding termite genera (*Mastotermes*, *Porotermes*, *Microcerotermes* and *Nasutitermes*) obtained from Chapter 2 and 3. In addition, we recovered four population genomes belonging to the Fibrobacteres and TG3 phyla via genome binning. Thus,

Chapter 5 characterises the functional features of Fibrobacteres, an important bacterial phylum in the termite gut ecosystem, providing insight into the evolutionary perspectives and classification of this novel phylum. Chapter 6 summarises findings of this thesis and suggests future directions for the study of termite gut microbiota.

1.6 References

Abubucker, S., Segata, N., Goll, J., Schubert, A.M., Izard, J., Cantarel, B.L. et al. (2012) Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Computational Biology* **8**: e1002358.

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., and Nielsen, P.H. (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology* **31**: 533-538.

Allwood, A.C., Walter, M.R., Kamber, B.S., Marshall, C.P., and Burch, I.W. (2006) Stromatolite reef from the Early Archaean era of Australia. *Nature* **441**: 714-718.

Amann, R.L., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiological Reviews* **59**: 143-169.

Aylward, F.O., Burnum, K.E., Scott, J.J., Suen, G., Tringe, S.G., Adams, S.M. et al. (2012) Metagenomic and metaproteomic insights into bacterial communities in leaf-cutter ant fungus gardens. *The ISME Journal* **6**: 1688-1701.

Bastien, G., Arnal, G., Bozonnet, S., Laguerre, S., Ferreira, F., Fauré, R. et al. (2013) Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics. *Biotechnology for Biofuels* **6**: 78.

Bauer, S., Tholen, A., Overmann, J., and Brune, A. (2000) Characterization of abundance and diversity of lactic acid bacteria in the hindgut of wood- and soil-feeding termites by molecular and culture-dependent techniques. *Archives of Microbiology* **173**: 126-137.

de la Bastide, M., and McCombie, W.R. (2002) Assembling genomic DNA sequences with PHRAP. In *Current Protocols in Bioinformatics*: John Wiley & Sons, Inc.

Beguin, P. (1990) Molecular biology of cellulose degradation. *Annual Review of Microbiology* **44**: 219-248.

Berchtold, M., Chatzinotas, A., Schönhuber, W., Brune, A., Amann, R., Hahn, D., and König, H. (1999) Differential enumeration and *in situ* localization of microorganisms in the hindgut of the lower termite *Mastotermes darwiniensis* by hybridization with rRNA-targeted probes. *Archives of Microbiology* **172**: 407-416.

Berchtold, M., Ludwig, W., and König, H. (1994) 16S rDNA sequence and phylogenetic position of an uncultivated spirochete from the hindgut of the termite *Mastotermes darwiniensis* Froggatt. *FEMS Microbiology Letters* **123**: 269-273.

Berlanga, M., Paster, B.J., and Guerrero, R. (2009) The taxophysiological paradox: changes in the intestinal microbiota of the xylophagous cockroach *Cryptocercus punctulatus* depending on the physiological state of the host. *International Microbiology* **12**: 227-236.

- Besemer, J., and Borodovsky, M. (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research* **33**: 451-454.
- Bignell, D., Eggleton, P., Nunes, L., and Thomas, K. (1997) Termites as mediators of carbon fluxes in tropical forest: budgets for carbon dioxide and methane emissions. In *Forests and Insects*. Watt, A.D., Stork, N.E., and Hunter, M.D. (eds). Cambridge: Chapman & Hall, pp. 109-113.
- Blaha, C.A.G., Santos, I.L.V.d.L., Pereira, M.S., Carvalho, C.M., Da Silva, C.L.V., De Melo, J.T.A. et al. (2007) Metagenomic evaluation of indigenous microbial diversity in tropical mangrove and semi-arid soils from petroliferous basin. *Journal of Biotechnology* **131**: S239.
- Boisvert, S., Raymond, F., Godzaridis, É., Laviolette, F., and Corbeil, J. (2012) Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biology* **13**: R122.
- Bonfante, P., and Genre, A. (2010) Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nature Communications* **1**: 48.
- Bonfante, P., Visick, K., and Ohkuma, M. (2010) Symbiosis. *Environmental Microbiology Reports* **2**: 475-478.
- Boucias, D.G., Cai, Y., Sun, Y., Lietze, V.U., Sen, R., Raychoudhury, R., and Scharf, M.E. (2013) The hindgut lumen prokaryotic microbiota of the termite *Reticulitermes flavipes* and its responses to dietary lignocellulose composition. *Molecular Ecology* **22**: 1836-1853.
- Bourguignon, T., Lo, N., Cameron, S.L., Šobotník, J., Hayashi, Y., Shigenobu, S. et al. (2015) The evolutionary history of termites as inferred from 66 mitochondrial genomes. *Molecular Biology and Evolution* **32**: 406-421.
- Brady, A., and Salzberg, S.L. (2009) Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nature Methods* **6**: 673-676.
- Brandl, R., Hyodo, F., Korff-Schmising, M.v., Maekawa, K., Miura, T., Takematsu, Y. et al. (2007) Divergence times in the termite genus *Macrotermes* (Isoptera: Termitidae). *Molecular Phylogenetics and Evolution* **45**: 239-250.
- Brauman, A., Dore, J., Eggleton, P., Bignell, D., Breznak, J.A., and Kane, M.D. (2001) Molecular phylogenetic profiling of prokaryotic communities in guts of termites with different feeding habits. *FEMS Microbiology Ecology* **35**: 27-36.
- Breznak, J. (2004) Invertebrates-insects. *Bull AT (ed), Microbial Diversity and Bioprospecting ASM Press, Washington, DC*: 191-203.
- Breznak, J.A., and Brune, A. (1994) Role of microorganisms in the digestion of lignocellulose by termites. *Annual Review of Entomology* **39**: 453-487.
- Brune, A. (2009) Chapter 249 - Symbionts Aiding Digestion. In *Encyclopedia of Insects (Second Edition)*. Vincent, H.R., and Ring, T.C. (eds). San Diego: Academic Press, pp. 978-983.
- Brune, A. (2014) Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology* **12**: 168-180.
- Brune, A., and Friedrich, M. (2000) Microecology of the termite gut: structure and function on a microscale. *Current Opinion in Microbiology* **3**: 263-269.

- Brune, A., and Ohkuma, M. (2011) Role of the termite gut microbiota in symbiotic digestion. In *Biology of Termites: a Modern Synthesis*. Bignell, D.E., Roisin, Y., and Lo, N. (eds): Springer Netherlands, pp. 439-475.
- Brune, A., and Dietrich, C. (2015) The gut microbiota of termites: Digesting the diversity in the light of ecology and evolution. *Annual Review of Microbiology* **69**.
- Burnum, K.E., Callister, S.J., Nicora, C.D., Purvine, S.O., Hugenholtz, P., Warnecke, F. et al. (2011) Proteome insights into the symbiotic relationship between a captive colony of *Nasutitermes corniger* and its hindgut microbiome. *The ISME Journal* **5**: 161-164.
- Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research* **37**: D233-D238.
- Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J. et al. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences* **108**: 4516-4522.
- Cardenas, E., and Tiedje, J.M. (2008) New tools for discovering and characterizing microbial diversity. *Current Opinion in Biotechnology* **19**: 544-549.
- Caspi, R., Altman, T., Dale, J.M., Dreher, K., Fulcher, C.A., Gilham, F. et al. (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research* **38**: D473-D479.
- Chaturvedi, V., and Verma, P. (2013) An overview of key pretreatment processes employed for bioconversion of lignocellulosic biomass into biofuels and value added products. *3 Biotech* **3**: 415-431.
- Colman, D.R., Toolson, E.C., and Takacs-Vesbach, C.D. (2012) Do diet and taxonomy influence insect gut bacterial communities? *Molecular Ecology* **21**: 5124-5137.
- Constantino, R. (Last updated: September 2012) On-line Termite Database.
- Cox-Foster, D.L., Conlan, S., Holmes, E.C., Palacios, G., Evans, J.D., Moran, N.A. et al. (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318**: 283-287.
- Creffield, J.W. (1996) *Wood destroying insects, wood borers and termites*: CSIRO Publications.
- Dai, X., Zhu, Y., Luo, Y., Song, L., Liu, D., Liu, L. et al. (2012) Metagenomic insights into the fibrolytic microbiome in yak rumen. *PLoS One* **7**: e40430.
- Davis, R.B., Baldauf, S.L., and Mayhew, P.J. (2009) Eusociality and the success of the termites: insights from a supertree of dictyopteran families. *Journal of Evolutionary Biology* **22**: 1750-1761.
- de Bary, A. (1879) *Die Erscheinung der Symbiose: Vortrag*.
- Decker, S.R., Siika-Aho, M., and Viikari, L. (2009) Enzymatic depolymerization of plant cell wall hemicelluloses. In *Biomass Recalcitrance*: Blackwell Publishing Ltd., pp. 352-373.
- DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.-U. et al. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496-503.

- Do, T.H., Nguyen, T.T., Nguyen, T.N., Le, Q.G., Nguyen, C., Kimura, K., and Truong, N.H. (2014) Mining biomass-degrading genes through Illumina-based de novo sequencing and metagenomic analysis of free-living bacteria in the gut of the lower termite *Coptotermes gestroi* harvested in Vietnam. *Journal of Bioscience and Bioengineering* **118**: 665-671.
- Douglas, A. (2009) The microbial dimension in insect nutritional ecology. *Functional Ecology* **23**: 38-47.
- Douglas, A.E. (2011) Lessons from studying insect symbioses. *Cell Host Microbe* **10**: 359-367.
- Dreyfus, M., and Régnier, P. (2002) The poly (A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. *Cell* **111**: 611-613.
- Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460-2461.
- Engel, M.S., Grimaldi, D.A., and Krishna, K. (2009) Termites (Isoptera): their phylogeny, classification, and rise to ecological dominance. *American Museum Novitates*: 1-27.
- Engel, P., and Moran, N.A. (2013) The gut microbiota of insects—diversity in structure and function. *FEMS Microbiology Reviews* **37**: 699-735.
- Engel, P., Martinson, V.G., and Moran, N.A. (2012) Functional diversity within the simple gut microbiota of the honey bee. *Proceedings of the National Academy of Sciences* **109**: 11002-11007.
- Engelbrekton, A., Kunin, V., Wrighton, K.C., Zvenigorodsky, N., Chen, F., Ochman, H., and Hugenholtz, P. (2010) Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *International Society for Microbial Ecology Journal* **4**: 642-647.
- Eutick, M.L., O'Brien, R.W., and Slaytor, M. (1976) Aerobic state of gut of *Nasutitermes exitiosus* and *Coptotermes lacteus*, high and low caste termites. *Journal of Insect Physiology* **22**: 1377-1380.
- Eutick, M.L., O'Brien, R.W., and Slaytor, M. (1978) Bacteria from the gut of Australian termites. *Applied and Environmental Microbiology* **35**: 823-828.
- Fall, S., Hamelin, J., Ndiaye, F., Assigbetse, K., Aragno, M., Chotte, J.L., and Brauman, A. (2007) Differences between bacterial communities in the gut of a soil-feeding termite (*Cubitermes niokoloensis*) and its mounds. *Applied and Environmental Microbiology* **73**: 5199-5208.
- Finn, R.D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R.Y., Eddy, S.R. et al. (2014) Pfam: the protein families database. *Nucleic Acids Research* **42**: D222-D230.
- Fisher, M., Miller, D., Brewster, C., Husseneder, C., and Dickerman, A. (2007) Diversity of gut bacteria of *Reticulitermes flavipes* as examined by 16S rRNA gene sequencing and amplified rDNA restriction analysis. *Current Microbiology* **55**: 254-259.
- Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Research* **32**: W273-W279.
- Gao, Q., Tancredi, S., and Thompson, G. (2012) Identification of mycosis-related genes in the eastern subterranean termite by suppression subtractive hybridization. *Archives of Insect Biochemistry and Physiology* **80**: 63-76.

- Geib, S.M., Filley, T.R., Hatcher, P.G., Hoover, K., Carlson, J.E., del Mar Jimenez-Gasco, M. et al. (2008) Lignin degradation in wood-feeding insects. *Proceedings of the National Academy of Sciences* **105**: 12932-12937.
- Graber, J.R., and Breznak, J.A. (2004) Physiology and nutrition of *Treponema primitia*, an H₂/CO₂-acetogenic spirochete from termite hindguts. *Applied and Environmental Microbiology* **70**: 1307-1314.
- Grieco, M.A.B., Cavalcante, J.J., Cardoso, A.M., Vieira, R.P., Machado, E.A., Clementino, M.M. et al. (2013) Microbial Community Diversity in the Gut of the South American Termite *Cornitermes cumulans* (Isoptera: Termitidae). *Microbial Ecology* **65**: 197-204.
- Guan, C., Ju, J., Borlee, B.R., Williamson, L.L., Shen, B., Raffa, K.F., and Handelsman, J. (2007) Signal mimics derived from a metagenomic analysis of the gypsy moth gut microbiota. *Applied and Environmental Microbiology* **73**: 3669-3676.
- Hadlington, P., and Staunton, I. (2008) *Australian Termites*. Sydney: University of New South Wales Press Ltd.
- Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews* **68**: 669-685.
- Hardison, R.C. (2003) Comparative genomics. *PLoS Biology* **1**: e58.
- He, S., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R.H. et al. (2013) Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS One* **8**: e61126.
- Herlemann, D.P., Geissinger, O., and Brune, A. (2007) The termite group I phylum is highly diverse and widespread in the environment. *Applied and Environmental Microbiology* **73**: 6682-6685.
- Hongoh, Y. (2010) Diversity and genomes of uncultured microbial symbionts in the termite gut. *Bioscience, Biotechnology, and Biochemistry* **74**: 1145-1151.
- Hongoh, Y., and Ishikawa, H. (1997) Uric acid as a nitrogen resource for the brown planthopper, *Nilaparvata lugens*: studies with synthetic diets and aposymbiotic insects. *Zoological science* **14**: 581-586.
- Hongoh, Y., Ohkuma, M., and Kudo, T. (2003) Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera; Rhinotermitidae). *FEMS Microbiology Ecology* **44**: 231-242.
- Hongoh, Y., Deevong, P., Hattori, S., Inoue, T., Noda, S., Noparatnaraporn, N. et al. (2006a) Phylogenetic diversity, localization, and cell morphologies of members of the candidate phylum TG3 and a subphylum in the phylum Fibrobacteres, recently discovered bacterial groups dominant in termite guts. *Applied and Environmental Microbiology* **72**: 6780-6788.
- Hongoh, Y., Ekpornprasit, L., Inoue, T., Moriya, S., Trakulnaleamsai, S., Ohkuma, M. et al. (2006b) Intracolony variation of bacterial gut microbiota among castes and ages in the fungus-growing termite *Macrotermes gilvus*. *Molecular Ecology* **15**: 505-516.
- Hongoh, Y., Deevong, P., Inoue, T., Moriya, S., Trakulnaleamsai, S., Ohkuma, M. et al. (2005) Intra- and interspecific comparisons of bacterial diversity and community structure support

- coevolution of gut microbiota and termite host. *Applied and Environmental Microbiology* **71**: 6590-6599.
- Huang, Q.-Y., Wang, W.-P., Mo, R.-Y., and Lei, C.-L. (2008) Studies on feeding and trophallaxis in the subterranean termite *Odontotermes formosanus* using rubidium chloride. *Entomologia Experimentalis et Applicata* **129**: 210-215.
- Hugenholtz, P., and Tyson, G.W. (2008) Microbiology: Metagenomics. *Nature* **455**: 481-483.
- Hugenholtz, P., Goebel, B.M., and Pace, N.R. (1998) Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity. *Journal of Bacteriology* **180**: 4765-4774.
- Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Research* **17**: 377-386.
- Hussain, A., Li, Y.-F., Cheng, Y., Liu, Y., Chen, C.-C., and Wen, S.-Y. (2013) Immune-related transcriptome of *Coptotermes formosanus* Shiraki workers: the defense mechanism. *PloS One* **8**: e69543.
- Husseneder, C., Ho, H.-Y., and Blackwell, M. (2010) Comparison of the bacterial symbiont composition of the Formosan subterranean termite from its native and introduced range. *The Open Microbiology Journal* **4**: 53.
- Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.
- Imelfort, M., Parks, D., Woodcroft, B.J., Dennis, P., Hugenholtz, P., and Tyson, G.W. (2014) GroopM: an automated tool for the recovery of population genomes from related metagenomes. *PeerJ* **2**: e603.
- Johjima, T., Taprab, Y., Noparatnaraporn, N., Kudo, T., and Ohkuma, M. (2006) Large-scale identification of transcripts expressed in a symbiotic fungus (*Termitomyces*) during plant biomass degradation. *Applied Microbiology and Biotechnology* **73**: 195-203.
- Kambhampati, S., and Eggleton, P. (2000) Termites: evolution, sociality, symbioses, ecology. In *Taxonomy and phylogeny of termites*. Abe, T., Bignell, D.E., and Higashi, M. (eds): Kluwer Academic, Dordrecht, The Netherlands.
- Kan, J., Hanson, T.E., Ginter, J.M., Wang, K., and Chen, F. (2005) Metaproteomic analysis of Chesapeake Bay microbial communities. *Saline Syst* **1**.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2011) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*: gkr988.
- Karp, P.D., and Caspi, R. (2011) A survey of metabolic databases emphasizing the MetaCyc family. *Archives of Toxicology* **85**: 1015-1033.
- Kaufman, M.G., Walker, E.D., Odelson, D.A., and Klug, M.J. (2000) Microbial community ecology insect nutrition. *American Entomologist* **46**: 173-185.
- Ke, J., Laskar, D.D., and Chen, S. (2013) Tetramethylammonium hydroxide (TMAH) thermochemolysis for probing *in situ* softwood lignin modification in each gut segment of the termite. *Journal of Agricultural and Food Chemistry* **61**: 1299-1308.

- Ke, J., Laskar, D.D., Singh, D., and Chen, S. (2011) *In situ* lignocellulosic unlocking mechanism for carbohydrate hydrolysis in termites: crucial lignin modification. *Biotechnology for Biofuels* **4**: 17.
- Kielbasa, S.M., Wan, R., Sato, K., Horton, P., and Frith, M.C. (2011) Adaptive seeds tame genomic sequence comparison. *Genome Research* **21**: 487-493.
- Köhler, T., Dietrich, C., Scheffrahn, R.H., and Brune, A. (2012) High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Applied and Environmental Microbiology* **78**: 4691-4701.
- Krause, L., Diaz, N.N., Goesmann, A., Kelley, S., Nattkemper, T.W., Rohwer, F. et al. (2008) Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Research* **36**: 2230-2239.
- Krishna, K., Grimaldi, D.A., Krishna, V., and Engel, M.S. (2013) Treatise on the isoptera of the world. *Bulletin of the American Museum of Natural History* **377**: 200-623.
- Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* **12**: 118-123.
- Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., and Hugenholtz, P. (2008) A bioinformatician's guide to metagenomics. *Microbiology and Molecular Biology Reviews* **72**: 557-578.
- Lax, A.R., and Osbrink, W.L.A. (2003) United States Department of Agriculture—Agriculture research service research on targeted management of the formosan subterranean termite *Coptotermes formosanus* Shiraki (Isoptera: Rhinotermitidae). *Pest Management Science* **59**: 788-800.
- Leadbetter, J., and Breznak, J. (1996) Physiological ecology of *Methanobrevibacter cuticularis* sp. nov. and *Methanobrevibacter curvatus* sp. nov., isolated from the hindgut of the termite *Reticulitermes flavipes*. *Applied and Environmental Microbiology* **62**: 3620-3631.
- Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R., and Gordon, J.I. (2008) Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology* **6**: 776-788.
- Li, Z., Chen, Y., Mu, D., Yuan, J., Shi, Y., Zhang, H. et al. (2012) Comparison of the two major classes of assembly algorithms: overlap–layout–consensus and de-bruijn-graph. *Briefings in Functional Genomics* **11**: 25-37.
- Lilburn, T., Schmidt, T., and Breznak, J. (1999) Phylogenetic diversity of termite gut spirochaetes. *Environmental Microbiology* **1**: 331-345.
- Liu, N., Yan, X., Zhang, M., Xie, L., Wang, Q., Huang, Y. et al. (2011) Microbiome of fungus-growing termites: a new reservoir for lignocellulase genes. *Applied and Environmental Microbiology* **77**: 48-56.
- Liu, N., Zhang, L., Zhou, H., Zhang, M., Yan, X., Wang, Q. et al. (2013) Metagenomic insights into metabolic capacities of the gut microbiota in a fungus-cultivating termite (*Odontotermes yunnanensis*). *PLoS One* **8**: e69184.

- Lo, N., Tokuda, G., Watanabe, H., Rose, H., Slaytor, M., Maekawa, K. et al. (2000) Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. *Current Biology* **10**: 801-804.
- Loman, N.J., Misra, R.V., Dallman, T.J., Constantinidou, C., Gharbia, S.E., Wain, J., and Pallen, M.J. (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nature biotechnology* **30**: 434-439.
- Makonde, H.M., Boga, H.I., Osiemo, Z., Mwirichia, R., Mackenzie, L.M., Göker, M., and Klenk, H.-P. (2013) 16S-rRNA-based analysis of bacterial diversity in the gut of fungus-cultivating termites (*Microtermes* and *Odontotermes* species). *Antonie Van Leeuwenhoek* **104**: 869-883.
- Mande, S.S., Mohammed, M.H., and Ghosh, T.S. (2012) Classification of metagenomic sequences: methods and challenges. *Briefings in Bioinformatics*: bbs054.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bembgen, L.A. et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.
- Margulis, L. (1981) *Symbiosis in cell evolution*. Biology Department, University of Massachusetts. Amherst. MA.
- Markowitz, V.M., Chen, I.-M.A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M. et al. (2014) IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Research* **42**: D568-D573.
- Mattéotti, C., Haubruge, E., Thonart, P., Francis, F., De Pauw, E., Portetelle, D., and Vandenberghe, M. (2011) Characterization of a new β -glucosidase/ β -xylosidase from the gut microbiota of the termite (*Reticulitermes santonensis*). *FEMS Microbiology Letters* **314**: 147-157.
- Mattéotti, C., Bauwens, J., Brasseur, C., Tarayre, C., Thonart, P., Destain, J. et al. (2012) Identification and characterization of a new xylanase from Gram-positive bacteria isolated from termite gut (*Reticulitermes santonensis*). *Protein Expression and Purification* **83**: 117-127.
- McDonald, J.E., Lockhart, R.J., Cox, M.J., Allison, H.E., and McCarthy, A.J. (2008) Detection of novel *Fibrobacter* populations in landfill sites and determination of their relative abundance via quantitative PCR. *Environmental Microbiology* **10**: 1310-1319.
- McFall-Ngai, M., Hadfield, M.G., Bosch, T.C., Carey, H.V., Domazet-Lošo, T., Douglas, A.E. et al. (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences* **110**: 3229-3236.
- McHardy, A.C., Martín, H.G., Tsirigos, A., Hugenholtz, P., and Rigoutsos, I. (2007) Accurate phylogenetic classification of variable-length DNA fragments. *Nature Methods* **4**: 63-72.
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M. et al. (2008) The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.
- Miller, J.R., Koren, S., and Sutton, G. (2010) Assembly algorithms for next-generation sequencing data. *Genomics* **95**: 315-327.
- Moran, N.A. (2001) Bacterial menageries inside insects. *Proceedings of the National Academy of Sciences* **98**: 1338-1340.

- Moran, N.A. (2002) Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**: 583-586.
- Moran, N.A. (2006) Symbiosis. *Current Biology* **16**: R866-871.
- Moran, N.A. (2007) Symbiosis as an adaptive process and source of phenotypic complexity. *Proceedings of the National Academy of Sciences* **104**: 8627-8633.
- Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L. et al. (2011) Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**: 970-974.
- Mullins, D.E., and Cochran, D.G. (1975) Nitrogen metabolism in the American cockroach—II. An examination of negative nitrogen balance with respect to mobilization of uric acid stores. *Comparative Biochemistry and Physiology Part A: Physiology* **50**: 501-510.
- Muyzer, G., and Smalla, K. (1998) Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie van Leeuwenhoek* **73**: 127-141.
- Namiki, T., Hachiya, T., Tanaka, H., and Sakakibara, Y. (2012) MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Research* **40**: e155-e155.
- Nimchua, T., Uengwetwanit, T., and Eurwilaichitr, L. (2012) Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes. *Journal of Microbiology and Biotechnology* **22**: 462-469.
- Noda, S., Kitade, O., Inoue, T., Kawai, M., Kanuka, M., Hiroshima, K. et al. (2007) Cospeciation in the triplex symbiosis of termite gut protists (*Pseudotriconympha* spp.), their hosts, and their bacterial endosymbionts. *Molecular Ecology* **16**: 1257-1266.
- Noguchi, H., Park, J., and Takagi, T. (2006) MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Research* **34**: 5623-5630.
- O'Brien, R., and Slaytor, M. (1982) Role of Microorganisms in the metabolism of termites. *Australian Journal of Biological Sciences* **35**: 239-262.
- Ohkuma, M. (2003) Termite symbiotic systems: efficient bio-recycling of lignocellulose. *Applied Microbiology and Biotechnology* **61**: 1-9.
- Ohkuma, M., and Kudo, T. (1996) Phylogenetic diversity of the intestinal bacterial community in the termite *Reticulitermes speratus*. *Applied and Environmental Microbiology* **62**: 461-468.
- Ohkuma, M., Noda, S., and Kudo, T. (1999) Phylogenetic diversity of nitrogen fixation genes in the symbiotic microbial community in the gut of diverse termites. *Applied and Environmental Microbiology* **65**: 4926-4934.
- Ohkuma, M., Noda, S., Hongoh, Y., Nalepa, C.A., and Inoue, T. (2009) Inheritance and diversification of symbiotic trichonymphid flagellates from a common ancestor of termites and the cockroach *Cryptocercus*. *Proceedings of the Royal Society of London B: Biological Sciences* **276**: 239-245.

- Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M.H.H., Oksanen, M.J., and Suggests, M. (2007) The vegan package. *Community Ecology Package*: 631-637.
- Otani, S., Mikaelyan, A., Nobre, T., Hansen, L.H., Koné, N.G.A., Sørensen, S.J. et al. (2014) Identifying the core microbial community in the gut of fungus-growing termites. *Molecular Ecology* **23**: 4631-4644.
- Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.-Y., Cohoon, M. et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research* **33**: 5691-5702.
- Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere. *Science* **276**: 734-740.
- Pace, N.R., Olsen, G.J., and Woese, C.R. (1986) Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell* **45**: 325-326.
- Parks, D.H., Tyson, G.W., Hugenholtz, P., and Beiko, R.G. (2014) STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**: 3123-3124.
- Peng, Y., Leung, H.C., Yiu, S.-M., and Chin, F.Y. (2011) Meta-IDBA: a de novo assembler for metagenomic data. *Bioinformatics* **27**: i94-i101.
- Peters, B., King, J., and Wylie, R. (1996) *Pests of Timber in Queensland*. Brisbane: Department of Primary Industries.
- Pontes, D.S., Lima-Bittencourt, C.I., Chartone-Souza, E., and Nascimento, A.M.A. (2007) Molecular approaches: advantages and artifacts in assessing bacterial diversity. *Journal of Industrial Microbiology and Biotechnology* **34**: 463-473.
- Poulsen, M., Hu, H., Li, C., Chen, Z., Xu, L., Otani, S. et al. (2014) Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proceedings of the National Academy of Sciences* **111**: 14500-14505.
- Rabiner, L.R., and Juang, B.-H. (1986) An introduction to hidden Markov models. *ASSP Magazine, IEEE* **3**: 4-16.
- Rashamuse, K., Mabizela-Mokoena, N., Sanyika, T.W., Mabvakure, B., and Brady, D. (2012) Accessing carboxylesterase diversity from termite hindgut symbionts through metagenomics. *Journal of Molecular Microbiology and Biotechnology* **22**: 277-286.
- Rashamuse, K., Ronneburg, T., Sanyika, W., Mathiba, K., Mmutlane, E., and Brady, D. (2014) Metagenomic mining of feruloyl esterases from termite enteric flora. *Applied Microbiology and Biotechnology* **98**: 727-737.
- Raychoudhury, R., Sen, R., Cai, Y., Sun, Y., Lietze, V.U., Boucias, D., and Scharf, M. (2013) Comparative metatranscriptomic signatures of wood and paper feeding in the gut of the termite *Reticulitermes flavipes* (Isoptera: Rhinotermitidae). *Insect Molecular Biology* **22(2)**: 155-171.
- Robinson, C.J., Bohannan, B.J.M., and Young, V.B. (2010) From structure to function: the ecology of host-associated microbial communities. *Microbiology and Molecular Biology Reviews* **74**: 453-476.

- Roisin, Y. (2000) Termites: Evolution, Sociality, Symbioses, Ecology. In *Diversity and evolution of caste patterns*. Abe, T., Bignell, D.E., and Higashi, M. (eds): Kluwer Academic, Dordrecht, The Netherlands.
- Rosen, G.L., Reichenberger, E.R., and Rosenfeld, A.M. (2011) NBC: the Naive Bayes Classification tool webserver for taxonomic classification of metagenomic reads. *Bioinformatics* **27**: 127-129.
- Rosengaus, R.B., Zecher, C.N., Schultheis, K.F., Brucker, R.M., and Bordenstein, S.R. (2011) Disruption of the termite gut microbiota and its prolonged consequences for fitness. *Applied and Environmental Microbiology* **77**: 4303-4312.
- Rosenthal, A.Z., Matson, E.G., Eldar, A., and Leadbetter, J.R. (2011) RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture. *The ISME Journal* **5**: 1133-1142.
- Rubin, E.M. (2008) Genomics of cellulosic biofuels. *Nature* **454**: 841-845.
- Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S. et al. (2007) The *Sorcerer II* global ocean sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biology* **5**: e77.
- Salmassi, T.M., and Leadbetter, J.R. (2003) Analysis of genes of tetrahydrofolate-dependent metabolism from cultivated spirochaetes and the gut community of the termite *Zootermopsis angusticollis*. *Microbiology* **149**: 2529-2537.
- Sanderson, K. (2011) Lignocellulose: A chewy problem. *Nature* **474**: S12-S14.
- Santo Domingo, J.W. (1998) Use of 16S rDNA community fingerprints to study cricket hindgut microbial communities. *The Journal of General and Applied Microbiology* **44**: 119-127.
- Sanyika, T.W., Rashamuse, K.J., Hennessy, F., and Brady, D. (2012) Luminal hindgut bacterial diversities of the grass and sugarcane feeding termite *Trinervitermes trinervoides*. *African Journal of Microbiology Research* **6**: 2639-2648.
- Scharf, M., Wu-Scharf, D., Zhou, X., Pittendrigh, B., and Bennett, G. (2005) Gene expression profiles among immature and adult reproductive castes of the termite *Reticulitermes flavipes*. *Insect Molecular Biology* **14**: 31-44.
- Scharf, M.E. (2015a) Termites as targets and models for biotechnology. *Annual Review of Entomology* **60**: 77-102.
- Scharf, M.E. (2015b) Omic research in termites: an overview and a roadmap. *Frontiers in Genetics* **6**.
- Scharf, M.E., Wu-Scharf, D., Pittendrigh, B.R., and Bennett, G.W. (2003) Caste-and development-associated gene expression in a lower termite. *Genome Biology* **4**: R62.
- Schmitt-Wagner, D., Friedrich, M.W., Wagner, B., and Brune, A. (2003) Phylogenetic diversity, abundance, and axial distribution of bacteria in the intestinal tract of two soil-feeding termites (*Cubitermes* spp.). *Applied and Environmental Microbiology* **69**: 6007-6017.
- Schultz, J.E., and Breznak, J.A. (1978) Heterotrophic bacteria present in hindguts of wood-eating termites [*Reticulitermes flavipes* (Kollar)]. *Applied and Environmental Microbiology* **35**: 930-936.

- Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30.14**: 2068-2069.
- Sen, R., Raychoudhury, R., Cai, Y., Sun, Y., Lietze, V.-U., Boucias, D.G., and Scharf, M.E. (2013) Differential impacts of juvenile hormone, soldier head extract and alternate caste phenotypes on host and symbiont transcriptome composition in the gut of the termite *Reticulitermes flavipes*. *BMC Genomics* **14**: 491.
- Sethi, A., Slack, J., Kovaleva, E.S., Buchman, G.W., and Scharf, M.E. (2012) Lignin-associated metagene expression in a lignocellulose-digesting termite. *Insect Biochemistry and Molecular Biology*.
- Sharon, I., Morowitz, M.J., Thomas, B.C., Costello, E.K., Relman, D.A., and Banfield, J.F. (2013) Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Research* **23**: 111-120.
- Sharpston, T.J. (2014) An introduction to the analysis of shotgun metagenomic data. *Frontiers in Plant Science* **5**.
- Shi, W., Syrenne, R., Sun, J.-Z., and Yuan, J.S. (2010) Molecular approaches to study the insect gut symbiotic microbiota at the 'omics' age. *Insect Science* **17**: 199-219.
- Simon, C., and Daniel, R. (2011) Metagenomic analyses: past and future trends. *Applied and Environmental Microbiology* **77**: 1153-1161.
- Smriga, S., Sandin, S.A., and Azam, F. (2010) Abundance, diversity, and activity of microbial assemblages associated with coral reef fish guts and feces. *FEMS Microbiology Ecology* **73**: 31-42.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R. et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America* **103**: 12115-12120.
- Sommer, D.D., Delcher, A.L., Salzberg, S.L., and Pop, M. (2007) Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* **8**: 64.
- St John, J. (2013) SeqPrep. In.
- Staley, J.T., and Konopka, A. (1985) Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology* **39**: 321-346.
- Steller, M.M., Kambhampati, S., and Caragea, D. (2010) Comparative analysis of expressed sequence tags from three castes and two life stages of the termite *Reticulitermes flavipes*. *BMC genomics* **11**: 463.
- Stingl, U., Radek, R., Yang, H., and Brune, A. (2005) "Endomicrobia": cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Applied and Environmental Microbiology* **71**: 1473-1479.
- Strassert, J.F., Desai, M.S., Radek, R., and Brune, A. (2010) Identification and localization of the multiple bacterial symbionts of the termite gut flagellate *Joenia annectens*. *Microbiology* **156**: 2068-2079.
- Strous, M., Kraft, B., Bisdorf, R., and Tegetmeyer, H.E. (2012) The binning of metagenomic contigs for microbial physiology of mixed cultures. *Frontiers in microbiology* **3**.

- Su, C., Lei, L., Duan, Y., Zhang, K.-Q., and Yang, J. (2012) Culture-independent methods for studying environmental microorganisms: methods, application, and perspective. *Applied Microbiology and Biotechnology* **93**: 993-1003.
- Suen, G., Weimer, P.J., Stevenson, D.M., Aylward, F.O., Boyum, J., Deneke, J. et al. (2011) The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. *PloS One* **6**: e18814.
- Tai, V., and Keeling, P.J. (2013) Termite hindguts and the ecology of microbial communities in the sequencing age. *Journal of Eukaryotic Microbiology* **60**: 421-428.
- Tai, V., James, E.R., Perlman, S.J., and Keeling, P.J. (2013) Single-cell DNA barcoding using sequences from the small subunit rRNA and internal transcribed spacer region identifies new species of *Trichonympha* and *Trichomitopsis* from the hindgut of the termite *Zootermopsis angusticollis*. *PloS One* **8**: e58728.
- Tai, V., James, E.R., Nalepa, C.A., Scheffrahn, R.H., Perlman, S.J., and Keeling, P.J. (2015) The role of host phylogeny varies in shaping microbial diversity in the hindguts of lower termites. *Applied and Environmental Microbiology* **81**: 1059-1070.
- Tartar, A., Wheeler, M., Zhou, X., Coy, M., Boucias, D., and Scharf, M. (2009) Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnology for Biofuels* **2**: 25.
- Tholen, A., Schink, B., and Brune, A. (1997) The gut microflora of *Reticulitermes flavipes*, its relation to oxygen, and evidence for oxygen-dependent acetogenesis by the most abundant *Enterococcus* sp. *FEMS Microbiology Ecology* **24**: 137-149.
- Thomas, T., Gilbert, J., and Meyer, F. (2012) Metagenomics-a guide from sampling to data analysis. *Microbial Informatics and Experimentation* **2**: 1-12.
- Thorne, B.L. (1997) Evolution of eusociality in termites. *Annual Review of Ecology and Systematics*: 27-54.
- Todaka, N., Inoue, T., Saita, K., Ohkuma, M., Nalepa, C.A., Lenz, M. et al. (2010) Phylogenetic analysis of cellulolytic enzyme genes from representative lineages of termites and a related cockroach. *PloS One* **5**: e8636.
- Todaka, N., Moriya, S., Saita, K., Hondo, T., Kiuchi, I., Takasu, H. et al. (2007) Environmental cDNA analysis of the genes involved in lignocellulose digestion in the symbiotic protist community of *Reticulitermes speratus*. *FEMS Microbiology Ecology* **59**: 592-599.
- Tokuda, G., Lo, N., Watanabe, H., Arakawa, G., Matsumoto, T., and Noda, H. (2004) Major alteration of the expression site of endogenous cellulases in members of an apical termite lineage. *Molecular Ecology* **13**: 3219-3228.
- Tokuda, G., and Watanabe, H. (2007) Hidden cellulases in termites: revision of an old hypothesis. *Biology Letters* **3**: 336-339.
- Tokuda, G., Tsuboi, Y., Kihara, K., Saitou, S., Moriya, S., Lo, N., and Kikuchi, J. (2014) Metabolomic profiling of ¹³C-labelled cellulose digestion in a lower termite: insights into gut symbiont function. *Proceedings of the Royal Society of London B: Biological Sciences* **281**: 20140990.

- Trimble, W.L., Keegan, K.P., D'Souza, M., Wilke, A., Wilkening, J., Gilbert, J., and Meyer, F. (2012) Short-read reading-frame predictors are not created equal: sequence error causes loss of signal. *BMC Bioinformatics* **13**: 183.
- Tringe, S.G., and Hugenholtz, P. (2008) A renaissance for the pioneering 16S rRNA gene. *Current Opinion in Microbiology* **11**: 442-446.
- Tringe, S.G., Von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W. et al. (2005) Comparative metagenomics of microbial communities. *Science* **308**: 554-557.
- Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M. et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37-43.
- Ultsch, A., and Mörchen, F. (2005) ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM.
- Van Valen, L. (1973) A new evolutionary law. *Evolutionary theory* **1**: 1-30.
- Vanwonterghem, I., Jensen, P.D., Ho, D.P., Batstone, D.J., and Tyson, G.W. (2014a) Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. *Current Opinion in Biotechnology* **27**: 55-64.
- Vanwonterghem, I., Jensen, P.D., Dennis, P.G., Hugenholtz, P., Rabaey, K., and Tyson, G.W. (2014b) Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. *The ISME Journal* **8**: 2015-2028.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A. et al. (2004) Environmental genome shotgun sequencing of the Sargasso sea. *Science* **304**: 66-74.
- Verma, M., Sharma, S., and Prasad, R. (2009) Biological alternatives for termite control: a review. *International Biodeterioration & Biodegradation* **63**: 959-972.
- Wang, Q., Qian, C., Zhang, X.-Z., Liu, N., Yan, X., and Zhou, Z. (2012) Characterization of a novel thermostable β -glucosidase from a metagenomic library of termite gut. *Enzyme and Microbial Technology* **51**: 319-324.
- Ward, N.L., Steven, B., Penn, K., Methé, B.A., and Detrich III, W.H. (2009) Characterization of the intestinal microbiota of two Antarctic notothenioid fish species. *Extremophiles* **13**: 679-685.
- Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T. et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**: 560-565.
- Watanabe, H., Noda, H., Tokuda, G., and Lo, N. (1998) A cellulase gene of termite origin. *Nature* **394**: 330-331.
- Weil, T., Rehli, M., and Korb, J. (2007) Molecular basis for the reproductive division of labour in a lower termite. *BMC Genomics* **8**: 198.
- Willner, D., and Hugenholtz, P. (2013) Metagenomics and community profiling: Culture-independent techniques in the clinical laboratory. *Clinical Microbiology Newsletter* **35**: 1-9.
- Woese, C.R. (1987) Bacterial evolution. *Microbiological Reviews* **51**: 221-271.

- Wood, T.G. and Johnson, R.A. (1986) The biology, physiology and ecology of termites. In: Economic impact and control of social insects, Vinson, S.B. (ed.). Praeger, New York, USA., pp. 1-68.
- Wu-Scharf, D., Scharf, M.E., Pittendrigh, B.R., and Bennett, G.W. (2003) 479 expressed Sequence Tags from a polyphenic *Reticulitermes flavipes* (Isoptera: Rhinotermitidae) cDNA library. *Sociobiology* **41**: 479-490.
- Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S.A. et al. (2011) Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**: 105-108.
- Wu, M., and Eisen, J.A. (2008) A simple, fast, and accurate method of phylogenomic inference. *Genome Biology* **9**: R151.
- Wyman, C., and Yang, B. (2009) Cellulosic biomass could help meet California's transportation fuel needs. *California Agriculture* **63**: 185-190.
- Xie, L., Zhang, L., Zhong, Y., Liu, N., Long, Y., Wang, S. et al. (2012) Profiling the metatranscriptome of the protistan community in *Coptotermes formosanus* with emphasis on the lignocellulolytic system. *Genomics* **99**: 246-255.
- Yang, F., Xu, B., Li, J., and Huang, Z. (2012) Transcriptome analysis of *Termitomyces albuminosus* reveals the biodegradation of lignocellulose. *Wei Sheng Wu Xue Bao* **52**: 466-477.
- Yang, H., Schmitt-Wagner, D., Stingl, U., and Brune, A. (2005) Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environmental Microbiology* **7**: 916-932.
- Yuki, M., Moriya, S., Inoue, T., and Kudo, T. (2008) Transcriptome analysis of the digestive organs of *Hodotermopsis sjostedti*, a lower termite that hosts mutualistic microorganisms in its hindgut. *Zoological Science* **25**: 401-406.
- Zhang, D., Lax, A., Henrissat, B., Coutinho, P., Katiya, N., Nierman, W.C., and Fedorova, N. (2012) Carbohydrate-active enzymes revealed in *Coptotermes formosanus* (Isoptera: Rhinotermitidae) transcriptome. *Insect Molecular Biology* **21**: 235-245.
- Zhao, Y., Tang, H., and Ye, Y. (2012) RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* **28**: 125-126.

Chapter 2 A molecular survey of Australian and North American termite genera indicates that vertical inheritance is the primary force shaping termite gut microbiomes

Nurdyana Abdul Rahman¹, Donovan H. Parks¹, Dana L. Willner^{1,2}, Anna L. Engelbrekton^{3,4}, Shana K. Goffredi⁵, Falk Warnecke^{3,6}, Rudolf H. Scheffrahn⁷ and Philip Hugenholtz^{1,3}

¹ Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Brisbane, Queensland, Australia; ² Current address: Department of Statistics, University of Illinois Urbana-Champaign, Champaign, IL, USA; ³ DOE Joint Genome Institute, Walnut Creek, CA, USA; ⁴ Current address: Energy Biosciences Institute, University of California, Berkeley, CA, USA; ⁵ Biology Department, Occidental College, Los Angeles, CA, USA; ⁶ Jena School for Microbial Communication (JSMC) and Microbial Ecology Group; ⁷ Fort Lauderdale Research and Education Center, University of Florida, Davie, Gainesville, FL, USA, Friedrich Schiller University Jena, Jena, Germany

Abstract

Termites and their microbial gut symbionts are major recyclers of lignocellulosic biomass. This important symbiosis is obligate but relatively open and more complex in comparison to other well-known insect symbioses such as the strict vertical transmission of *Buchnera* in aphids. The relative roles of vertical inheritance and environmental factors such as diet in shaping the termite gut microbiome are not well understood. The gut microbiomes of 66 specimens representing seven higher and nine lower termite genera collected in Australia and North America were profiled by small subunit (SSU) rRNA amplicon pyrosequencing. These represent the first reported culture-independent gut microbiome data for three higher termite genera: *Tenuirostritermes*, *Drepanotermes*, and *Gnathamitermes*; and two lower termite genera: *Marginitermes* and *Porotermes*. Consistent with previous studies, bacteria comprise the largest fraction of termite gut symbionts, of which 11 phylotypes (6 *Treponema*, 1 *Desulfarculus*-like, 1 *Desulfovibrio*, 1 *Anaerovorax*-like, 1 *Sporobacter*-like, and 1 *Pirellula*-like) were widespread occurring in $\geq 50\%$ of collected specimens. Archaea are generally considered to comprise only a minority of the termite gut microbiota (<3%); however, archaeal relative abundance was substantially higher and variable in a number of specimens including *Macrognathotermes*, *Coptotermes*, *Schedorhinotermes*, *Porotermes*, and *Mastotermes* (representing up to 54% of amplicon reads). A ciliate related to *Clevelandella* was detected in low abundance in *Gnathamitermes* indicating that protists were either reacquired after protists loss in higher termites or persisted in low numbers across this transition. Phylogenetic analyses of the bacterial communities indicate that vertical inheritance is the primary force shaping termite gut microbiota. The effect of diet is secondary and appears to influence the relative abundance, but not membership, of the gut communities. Vertical inheritance is the primary force shaping the termite gut microbiome indicating that species are successfully and faithfully passed from one generation to the next via trophallaxis or coprophagy. Changes in relative abundance can occur on shorter time scales and appear to be an adaptive mechanism for dietary fluctuations.

2.1 Introduction

Co-evolution of microbial species with eukaryotic hosts is well known for obligate endosymbionts such as *Buchnera* in aphids (Moran and Mira, 2001) and *Wolbachia* in nematodes (Bandi et al., 1998). The importance of vertical inheritance is less clear in more open symbioses such as the microbiota of gastrointestinal tracts in which environmental perturbations and lateral transfer of organisms between hosts may play a more prominent role. Using culture-independent small subunit (SSU) rRNA-based community profiling, Ley et al. (Ley et al., 2008b; Ley et al., 2008a) found that both host phylogeny and diet shape gut microbiomes in many mammalian species and Ochman et al. concluded that vertical inheritance of gut microbiota in primates is discernable over evolutionary time scales (Ochman et al., 2010).

Termites provide an appealing model system to explore the relative importance of vertical inheritance and environmental factors on symbiotic gut microbiota as unlike most insects, their gut communities are relatively complex comprising in the order of hundreds of species (Douglas, 2011). Termites are thought to have evolved from a cockroach-like ancestor into strictly eusocial insects that feed exclusively on lignocellulosic biomass (Brune, 2014). Such recalcitrant substrates are digested through an obligate symbiosis with specialised gut microbiota comprising bacteria and protists in lower termites (classified into eight families) and bacteria only, in more recently evolved higher termites (classified in a single family, the Termitidae) (Krishna et al., 2013). Accordingly, transmission of gut microorganisms between termites is more strictly regulated than in mammals via trophallaxis (mouth to mouth transmission) or coprophagy (consumption of faeces) (Noda et al., 2007) and co-speciation with the host has been observed in selected members of the gut community (Noda et al., 2007). To determine whether vertical inheritance is the dominant force shaping termite gut communities more broadly, we used SSU rRNA gene amplicon pyrosequencing to profile the gut microbiomes of 66 termite samples, representing 16 genera, obtained in Australia and North America. These data expand current knowledge of termite gut microbiome diversity and represent the first gut community profiles for three higher (*Tenuirostritermes*, *Drepanotermes*, *Gnathamitermes*) and two lower (*Marginitermes*, *Porotermes*) termite genera.

2.2 Methods

2.2.1 Sample collection and processing

Termite collections were made on public lands in Queensland, Northern Territory (Australia), and Arizona (United States of America). Where possible, specimens were collected with their nest material and transported to the laboratory in ventilated plastic containers at room temperature to

reduce stress to the insects. Termites were removed from their nest material within a day of arriving in the laboratory. For community profiling, workers were transferred to a metal tray and frozen at -80°C for 20 min, then collected into 2 ml cryotubes and stored at -20°C until further processing. Frozen specimens were thawed on ice, and gut tracts were extracted using clean sharp tweezers. The guts were immediately transferred into a sterile 1.5 ml microtube on ice and stored at -20°C until extraction. For morphological identification, soldier specimens were stored in 85% ethanol.

2.2.2 DNA extraction

Total genomic DNAs were extracted from pooled (5–30) whole gut samples, depending on size of species, using FastDNA® SPIN kit for Soil (MP Biomedicals, Australia). Termite guts were added to a lysing matrix, treated with lysis buffer, and underwent bead beating in the Vortex-Genie® 2 (MoBio Laboratories, USA). DNA was bound to silica matrix and washed and eluted in DNase-free water. DNA yield was then quantified by the Qubit™ fluorometer and QuantIT ds-DNA BR assay kit (Invitrogen, Australia). DNA concentration varied depending on the biomass of the whole gut. DNA concentrations were standardised across all samples to 20 $\mu\text{g}/\text{ml}$, diluting where necessary in Ultrapure™ distilled water (Invitrogen, Australia). DNA quality was evaluated using gel electrophoresis on 1.0% agarose gels stained with SYBR Safe, visualised on a CCD compact image system (Major Science, USA).

2.2.3 SSU rRNA PCR and amplicon pyrosequencing

The universal primer pair 926F (or prokaryote-specific 803F) and 1392R was used to amplify the V6 to V8 variable regions of the SSU rRNA gene. Primer sequences were modified by incorporation of the Roche 454 A or B adaptor sequences and a unique 5–7 nucleotide barcode, known as multiplex identifier (MID), to identify amplicons originating from different samples in the same sequencing reaction. The reverse primer 1392R was barcoded on the 5' end with the MID between the 454 A adaptor (uppercase) and the SSU rRNA primer (lowercase) (5'-CCA TCT CAT CCC TGC GTG TCT CCG AC TCAG [MID] acgggcggtgtgRc-3'); and the 926 forward primer (lowercase) was modified by addition of 454 B adaptor (uppercase) at its 5' end (3'-CCT ATC CCC TGT GTG CCT TGG CAG TC TCAG aaactYaaaKgaattgRcgg-3') (or 803 forward primer ttgaKacccBNgtagtc) (Engelbrektson et al., 2010).

DNA amplification was carried out in 50 μl PCR reactions, using 2 μl of termite whole gut DNA extract as template. The amplification mixture contained 0.2 μl of 1U Fisher BioReagents* *Taq* DNA polymerase (Thermo Fisher Scientific Inc., USA), 4 μl of 25 mM MgCl_2 , 1.5 μl of BSA (Roche diagnostic, Australia), 5 μl of 10X buffer, 1 μl of dNTP mix (each at a concentration of 10

mM), 1 µl of each 10 mM forward primer and reverse primer; and *E.coli* was used as the positive control. PCR was performed using a Veriti® Thermal Cycler (Applied Biosystems™, Australia) with the following cycling parameters: an initial denaturation step of 95°C for 3 min followed by 30 cycles of 95°C for 30 s, 55°C for 45 s, 72°C for 90 s, and a final extension of 72°C for 10 min. Amplification products were quantified by electrophoresis in 1% agarose with SYBR Safe staining.

To ensure that similar numbers of sequencing reads were obtained for each sample, PCR amplicons were pooled in equal concentrations after cycling and then purified using the Agencourt® AMPure® XP Kit (Beckman, USA). DNA was quantified with the Qubit™ fluorometer and QuantIT ds-DNA BR assay kit. Cleaned, pooled, barcoded amplicons were submitted for pyrosequencing library preparation where they were mixed in equal proportions with other samples prior to emulsion PCR for GS FLX pyrosequencing (454 Life Sciences, USA).

2.2.4 Analysis of SSU rRNA gene sequences

SSU rRNA sequence data were obtained from the multiplexed 454 run by converting the pyrosequencing flowgrams to sequence reads using the standard software provided by 454 Life Sciences (Engelbrektson et al., 2010; Kunin et al., 2010). Short and/or low quality reads were removed using UCHIME version 4.2 (Bragg et al., 2012), and homopolymer errors were corrected using Acacia (Bragg et al., 2012). Sequence data were analysed using a pyrotag (pyrosequence reads) processing pipeline, Quantitative Insights Into Microbial Ecology (QIIME) (Caporaso et al., 2010) and CD-HIT (Fu et al., 2012). Reads were hard trimmed to 250 bp and clustered into operational taxonomic units (OTUs) with a threshold of 97% sequence identity using MCL (Kunin and Hugenholtz, 2010). OTU representatives were compared to the Greengenes database (February 2011 release) for taxonomy assignment using BLAST (Altschul et al., 1990; McDonald et al., 2012). A table which lists the relative abundance of each OTU in each sample was generated and visualised as a heatmap. The relationship between the microbial communities in different samples was assessed using jackknifed UPGMA trees derived from the distance matrices obtained with the phylogeny-based unweighted and weighted Soergel beta-diversity measures implemented in Express Beta Diversity v1.04 (Parks and Beiko, 2013). The Soergel distance measures community relatedness based on phylogeny and either presence/absence (unweighted) or relative abundance (weighted) of OTUs (Lozupone and Knight, 2005; Lozupone et al., 2006; Lozupone et al., 2011). A comparative analysis of several phylogenetic beta-diversity measures resulted in the recommendation of the Soergel measure based in part on the unweighted variant being identical to unweighted UniFrac and the weighted variant being closely related to normalised, weighted UniFrac (Parks and Beiko, 2013). The relative abundance of different phyla within the higher and

lower termites was compared using Welch's *t*-tests with Šidák multiple test correction as implemented in STAMP (Parks and Beiko, 2010).

2.2.5 Molecular identification of termite host species

The mitochondrial cytochrome oxidase II (COII) gene was amplified with PCR using three sets of primers Fleu/Rlys (TCT AAT ATG GCA GAT TAG TGC/GAG ACC AGT ACT TGC TTT CAG TCA TC), COIIF-M13/COIIR-M13 (GTT TTC CCA GTC ACG ACG TTG TAC AGA TAA GTG CAT TGG ATT T/AGG AAA CAG CTA TGA CCA TGG TTT AAG AGA CCA GTA CTT G), and COIIFw-M13/COIIRw-M13 (GTT TTC CCA GTC ACG ACG TTG TAC AGA YWA GTG CAH TGG ATT T/AGG AAA CAG CTA TGA CCA TGG TTT AAG AGA CCA KTA CTT G). This gene is commonly used for identification of termite species (Ohkuma et al., 2004; Legendre et al., 2008). The amplification products were purified and directly sent for Sanger sequencing (Macrogen Inc., Korea). The sequences were manually trimmed and inspected using Geneious software (www.geneious.com). Reference COII nucleotide sequences were obtained from the National Center for Biotechnology Information server (<http://www.ncbi.nlm.nih.gov>) and aligned using Clustal W in ARB (Price et al., 2009), followed by manual checking and refinement of the automated alignment. Nucleotide and amino acid-based trees were constructed using a neighbour-joining method in ARB, and the topologies were compared. The COII tree was inferred using FastTree v2.1.3 (Price et al., 2009) with the generalised time-reversible model of nucleotide evolution.

2.2.6 Stable isotope analysis

Two to three individual termites for each sample were dried at 60°C, homogenised, and acidified to remove inorganic carbonate. The abundances of ^{13}C and ^{12}C were determined at the School of Biological Sciences, Washington State University using continuous-flow isotope ratio mass spectrometry with a Costech elemental analyser coupled to a Micromass Isoprime isotope ratio mass spectrometer (EA/IRMS). The isotope ratios of ^{13}C to ^{12}C are reported relative to Pee Dee Belemnite.

2.2.7 Nucleotide sequence accession numbers

All SSU rRNA sequence data obtained from this study have been deposited in GenBank under BioProject PRJNA248567. The COII termite host sequences are deposited under the accession numbers KJ907786–KJ907853.

2.3 Results

2.3.1 Sample collection and host identification

Samples of 66 termite colonies and two cockroaches were collected in Australia (Queensland and the Northern Territory) and the United States (Arizona) (Appendix A: **Table S2.1**). Termites were identified by sequencing and comparative analysis of their mitochondrial cytochrome oxidase II (COII) genes (Liu and Beckenbach, 1992) using the cockroaches as outgroup taxa. They were classified according to their closest identified phylogenetic neighbour in the public reference database (Appendix A: **Figure S2.1**) and also by soldier morphology (Appendix A: **Figure S2.2**). A total of 16 termite genera were sampled, seven higher and nine lower termites representing five of the nine recognised families (**Table 2.1**). The phylogenetic tree used to classify our samples (Appendix A: **Figure S2.1**) is consistent with previous inferences based on COII and other marker genes (Thompson et al., 2000; Legendre et al., 2008; Ware et al., 2010) with the following observations. The genus *Nasutitermes* is not monophyletic (Inward et al., 2007), clustering together with several other nasute genera (subfamily Nasutitermitinae) including *Tumulitermes*, *Hospitalitermes*, and specimens 7TT2 and 7TT3, morphologically identified as *Tenuirostritermes*. Similarly, *Amitermes* is not monophyletic, clustering together with *Gnathamitermes* and *Drepanotermes*, although it should be noted that internal groupings within the Termitidae are not well supported by bootstrap resampling. Specimens 8MH1 and 9MH1 are the first COII data for the genus *Marginitermes*, and these sequences are grouped with members of the family Kalotermitidae as predicted by morphological similarities (Scheffrahn and Postle, 2013). All other COII sequences obtained from the collected specimens, including cockroach outgroups, are grouped with reference sequences belonging to the expected genera predicted by morphology (Appendix A: **Figure S2.1**). We then used this host phylogeny as a reference to establish the degree of vertical inheritance occurring with resident gut microbiomes.

Table 2.1: Summary of the surveyed 66 termite whole gut samples according to host phylogeny (genus and family); sample location (country); and relative bacterial, archaeal, and protist abundances using universal primers (926F) and prokaryote primers (803F) in some instances (*see* text and Appendix A: **Figure S2.10**).

Termite genus	Family	Number of samples			Bacteria		Archaea		Protist
		AU	US	Total	%		%		%
<i>Higher</i>									
1 <i>Drepanotermes</i>	Termitidae	1	0	1	926F	803F	926F	803F	0.0
2 <i>Gnathamitermes</i>	Termitidae	0	8	8	97.5		2.5		0.0
3 <i>Amitermes</i>	Termitidae	2	8	10	99.7		0.1		0.2
4 <i>Nasutitermes</i>	Termitidae	2	8	10	98.2		1.7		0.0
5 <i>Tenuirostritermes</i>	Termitidae	7	1	8	97.5		2.5		0.0
6 <i>Microcerotermes</i>	Termitidae	0	2	2	99.8		0.2		0.0
7 <i>Macrognathotermes</i>	Termitidae	12	0	12	99.1		0.9		0.0
		1	0	1	77.9	63.5	22.1	36.5	0.0
	<i>Sub-total</i>	24	18	42					
<i>Lower</i>									
8 <i>Reticulitermes</i>	Rhinotermitidae	0	3	3	92.0	803F	926F	803F	7.8
9 <i>Heterotermes</i>	Rhinotermitidae	6	0	6	91.4		5.3		3.3
10 <i>Coptotermes</i>	Rhinotermitidae	3	0	3	66.5	78.6	33.4	21.4	0.1
11 <i>Schedorhinotermes</i>	Rhinotermitidae	3	0	3	82.3	72.9	17.2	27.1	0.5
12 <i>Marginitermes</i>	Kalotermitidae	0	2	2	98.5		0.0		0.2
13 <i>Incisitermes</i>	Kalotermitidae	0	1	1	97.6		0.0		1.5
14 <i>Glyptotermes</i>	Kalotermitidae	2	0	2	100.0		0.0		2.4
15 <i>Porotermes</i>	Stolotermitidae	1	0	1	42.1	50.1	57.7	49.9	0.0
16 <i>Mastotermes</i>	Mastotermitidae	3	0	3	82.2	92.4	17.4	7.6	0.4
	<i>Sub-total</i>	18	6	24					
	<i>Overall</i>	42	24	66					

2.3.2 Gut microbiome profiling

Whole guts were removed and pooled from 5 to 30 workers depending on the size of the species (Appendix A: **Table S2.1**). In the case of the two cockroach outgroups, the gut material of a single individual was used for subsequent analyses. Culture-independent microbial community profiles were determined via SSU rRNA gene amplicon pyrosequencing using the primers 926F and 1392R that broadly target all three domains of life (Lane, 1991). To evaluate the reproducibility of the profiles based on sets of pooled workers, we initially generated three biological replicates for four samples representing different termite genera. Clustering of samples by redundancy analysis (RDA) using Hellinger transformed data showed that the variation between the biological replicates of each subsampled genus was significantly less than the variation between termite genera (Appendix A: **Figure S2.3**). Based on these observations and to permit a broader survey, we generated only one pooled worker sample profile for each of the remaining 62 termite specimens. A total of 457,947 pyrosequence raw reads were produced from the 68 samples ranging from 600 to 10,000 per sample after removal of termite (or cockroach) host SSU rRNA gene sequences, which comprised from 3% to 55% of total reads for each sample. Specimens were randomly resampled to a depth of 600 reads, and rarefaction and diversity analysis suggested that this was adequate to describe the overall

diversity of the samples (Appendix A: **Figure S2.4**). The resampled data was normalised for SSU rRNA copy number variation using CopyRighter (Angly et al., 2014) which can vary by up to an order of magnitude between prokaryotic genera. However, the effect of copy number correction was relatively subtle for these datasets (Appendix A: **Table S2.2**). Overall, the majority of non-host amplicon reads from the whole gut samples were bacterial (95.4% on average in higher termites, 83.8% in lower) with smaller percentages of archaea (4.5% in higher, 14.4% in lower) and protists (0.1% in higher, 1.1% in lower) recovered (**Table 2.1**).

2.3.3 Bacterial profiles

To determine the evolutionary distribution and conservation of bacterial groups across the sampled termite host radiation, we performed a prevalence versus relative abundance analysis (Turnbaugh et al., 2010). Beginning at the broad taxonomic rank of phylum, all termite gut microbiomes were noted to comprise a core set (100% prevalence) of four bacterial phyla (Bacteroidetes, Firmicutes, Spirochaetes, and Proteobacteria) and an accessory set (<100% prevalence) of six bacterial phyla (Elusimicrobia, Fibrobacteres, Actinobacteria, Synergistetes, Planctomycetes, and Acidobacteria) using a relative abundance threshold of 1% in at least one sample (**Table 2.1**). Within the termite cohort, the core and accessory phyla showed pronounced differences in prevalence and relative abundances most notably between lower and higher termites. On average across the sampled genera, the Bacteroidetes are more abundant in lower than in higher termites, and the Spirochaetes, Acidobacteria, Fibrobacteres, and Synergistetes are more abundant in higher than lower termites (**Table 2.1** and Appendix A: **Figure S2.5**). We also observed that the Elusimicrobia are highly abundant in many lower termites while being nearly absent in all higher termites (Appendix A: **Figure S2.6**). These differences in relative abundance are mostly accounted for by a small number of genera in each of the phyla (*see below*). Additionally, we noted a secondary pattern associated with diet at the phylum level. Polyphagous termite genera (i.e. those comprising species with different diets) tended to show an increase in the relative abundance of Spirochaetes and Fibrobacteres and a decrease of Firmicutes on a wood relative to a grass diet (*Nasutitermes*) and on a grass relative to dung diet (*Gnathamitermes*) (Appendix A: **Figure S2.7**).

Table 2.2: Summary of core and accessory bacterial phyla in higher and lower termite gut communities present at >1% relative abundance in at least one sample.

Phylum	Higher		Lower		p-values
	Prevalence %	Relative abundance % (S.D)	Prevalence %	Relative abundance % (S.D)	
Bacteroidetes	100.0	6.3 (±5.0)	100.0	41.3 (±24.8)	***
Firmicutes	100.0	24.0 (±14.1)	100.0	19.1 (±11.6)	—
Spirochaetes	100.0	44.3 (±18.9)	100.0	13.2(±13.0)	***
Proteobacteria	100.0	5.5 (±2.7)	100.0	7.5 (±6.5)	—
Planctomycetes	100.0	4.3 (±4.4)	79.1	2.3 (±2.6)	—
Synergistetes	95.2	3.1 (±3.0)	95.8	1.0 (±0.6)	—
Actinobacteria	92.9	1.8 (±1.8)	87.5	2.3 (±2.1)	—
Acidobacteria	90.5	2.0 (±1.3)	45.8	<1 (±0.8)	***
Fibrobacteres	95.2	5.7 (±5.2)	12.5	<1 (±1.1)	***
Elusimicrobia	31.0	<1 (±0.2)	70.8	8.4 (±15.3)	—
Phylum in bold		*** p-value <0.05			
S.D Standard deviation		— p-value >0.05			

For the Bacteroidetes, the genus *Candidatus Azobacteroides* is highly represented in many of the lower termite specimens, and for the Elusimicrobia, members of the genus *Candidatus Endomicrobium* are similarly highly represented in several lower termite genera (**Figure 2.1**). For the Spirochaetes, the genus *Treponema* is highly represented in all of the higher termite genera; and for the Fibrobacteres, which were not detected in most of the lower termite samples, members of the classes *Chitinovibrionae* (TG3) and Fibrobacteres-2 were broadly represented in higher termite specimens (**Figure 2.1**). At increased phylogenetic resolution, several operational taxonomic units (OTUs) stood out either because they were abundant (>10% of bacterial reads) in one or a few termite genera and/or prevalent in the surveyed termites (present in >50% of specimens) (**Figure 2.2**). Four OTUs belonging to *Candidatus Azobacteroides* represent on average >10% of the reads from the guts of a number of lower termite genera and appear to have a co-evolutionary signal. For example, OTU5 is found in five of the six *Heterotermes* specimens that cluster together in the COII tree (Appendix A: **Figure S2.1**), with the phylogenetic outlier, *Heterotermes* BF01 containing a different *Candidatus Azobacteroides* OTU (OTU7; **Figure 2.2**). Similarly, three abundant *Candidatus Endomicrobium* OTUs likely representing separate species occur in different lower termite genera (*Porotermes*—OTU43, *Incisitermes*—OTU55, *Reticulitermes*—OTU24; **Figure 2.2** and Appendix A: **Figure S2.8**). Other abundant OTUs included *Candidatus Vestibaculum* in *Incisitermes* (OTU27) and *Marginitermes* (OTU105), *Blattabacterium* in *Mastotermes* (OTU22) and in the cockroach outgroups (OTU3), *Enterococcus* (OTU44) in one *Coptotermes* sample (AP01), *Dysgonomonas* (OTU207) in one *Heterotermes* sample (SL01), and *Fusobacterium* (OTU133) in all three *Mastotermes* specimens. In terms of prevalence, *Treponema* was the standout

genus, with six *Treponema* OTUs being broadly represented across the higher termites and in some instances also across the lower termites, for example OTU1 (present in 92% of all specimens; **Figure 2.2**). To confirm that the ubiquity of this OTU was not due to sample contamination, we examined it at higher resolution by dividing the 7,223 reads comprising OTU1 into identical clusters (Appendix A: **Table S2.3**). Most (89%) of these identical clusters were from members of the same termite families suggesting minimal contamination (and vertical inheritance) and also indicating that while 97% OTUs reduce the effect of pyrosequencing error on diversity estimates (Kunin et al., 2010), they are often composites of multiple strains (Patin et al., 2013). Although OTU1 was present as a low abundance member in most termite genera (<1%), it was highly represented in *Microcerotermes* (up to 35% of bacterial reads; Appendix A: **Figure S2.7**). Other high prevalence (and mostly low abundance) OTUs included *Desulfarculus*-like (OTU51), *Desulfovibrio* (OTU38), *Anaerovorax*-like (OTU120), *Sporobacter*-like (OTU364), and *Pirellula*-like (OTU151) bacteria (**Figure 2.2** and Appendix A: **Figure S2.8**).

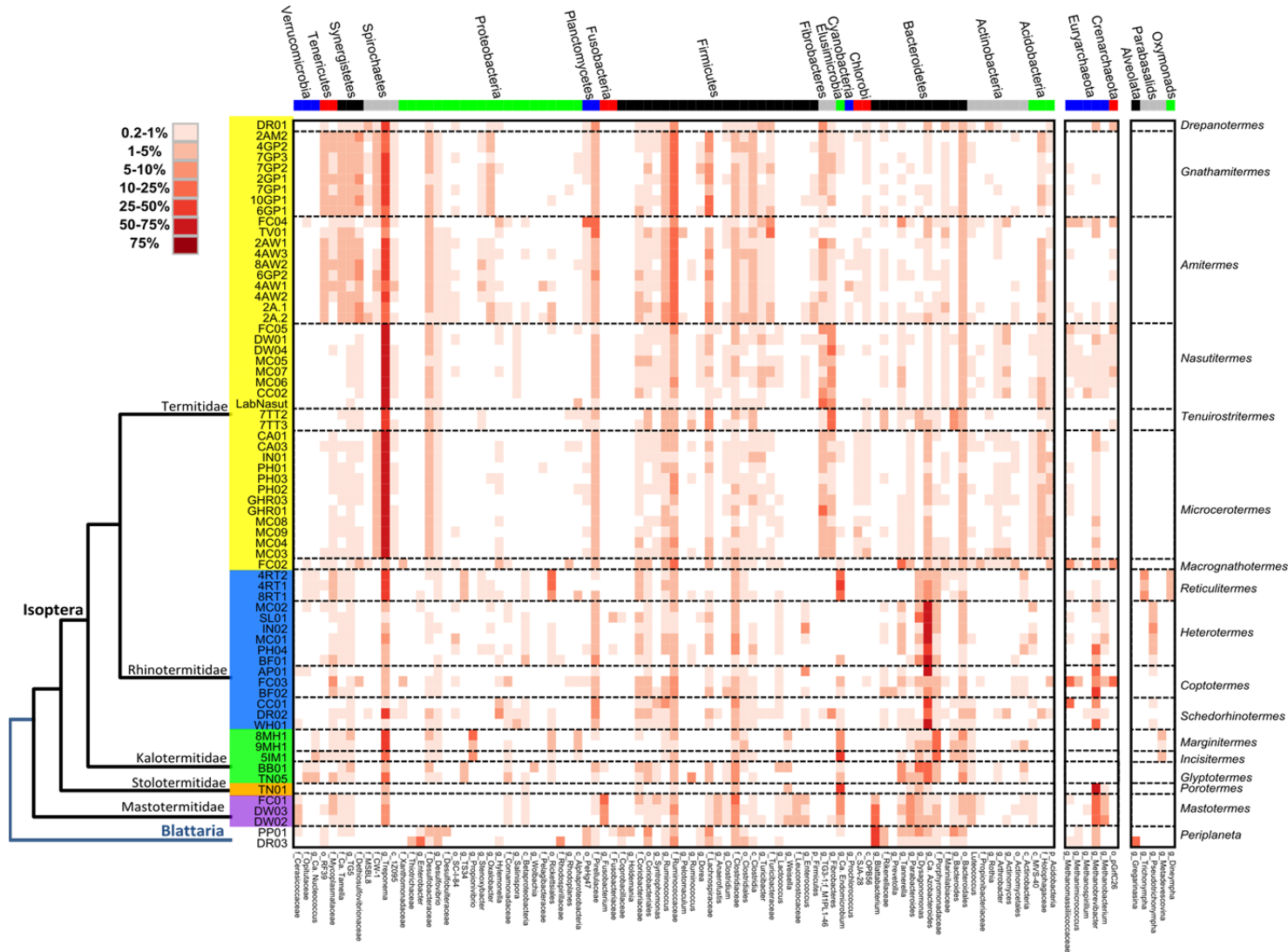


Figure 2.1: Heatmap showing microbial taxa (mostly genus and family level) with relative abundance $\geq 0.2\%$ in one or more whole gut samples surveyed in this study. Each row represents a gut sample and each column a microbial taxon with relative abundance indicated by shading according to the legend. Phylum-level designations for the microbial taxa are indicated at the top of the figure, and host sample phylogeny is indicated to the left (family) and right (genus) of the figure.

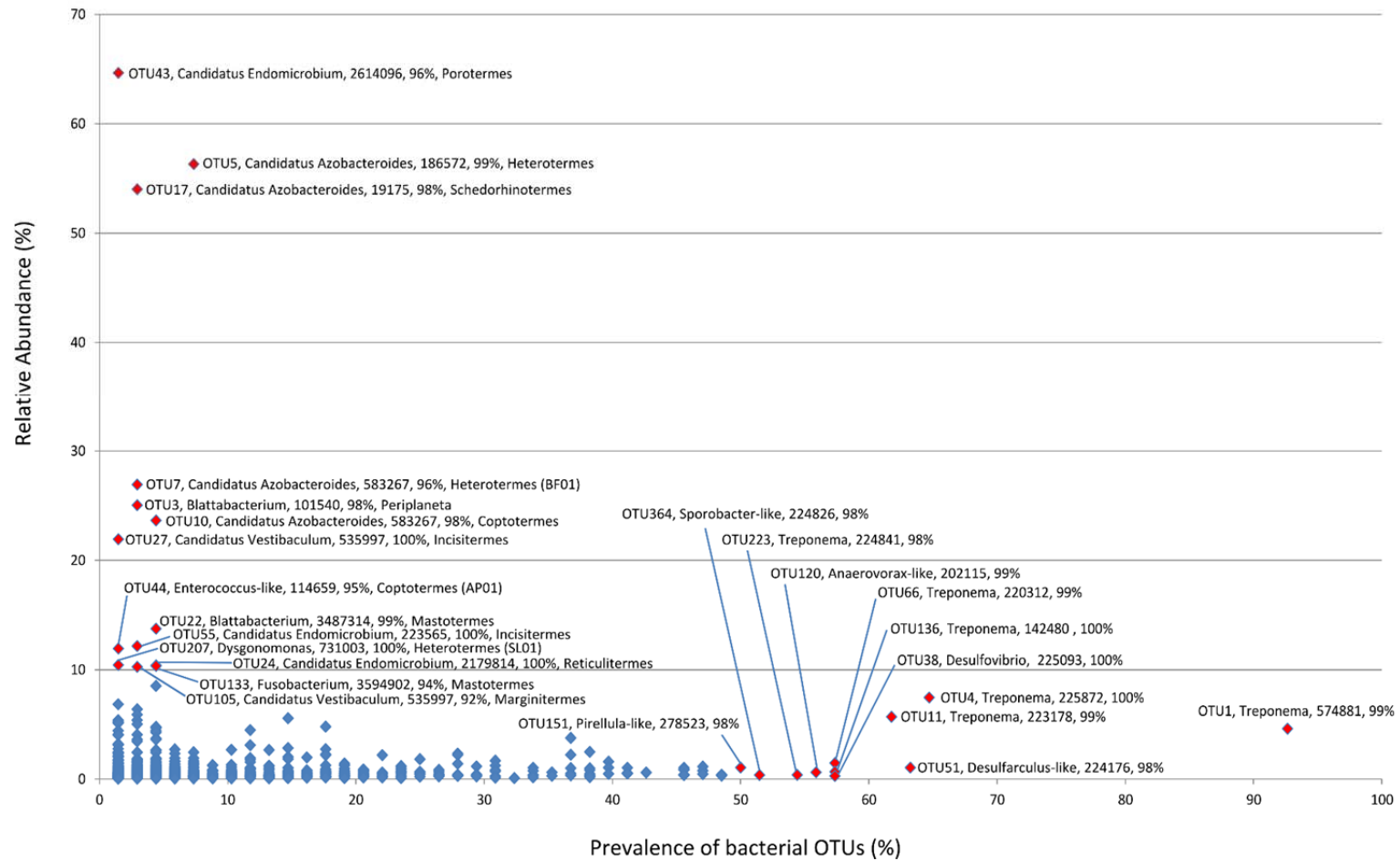


Figure 2.2: Prevalence versus relative abundance graph of bacterial OTUs (97% sequence identity) in the surveyed gut samples. OTUs with $\geq 10\%$ relative abundance or $\geq 50\%$ prevalence across the 66 termite samples are highlighted in red and labelled with OTU ID and closest matching bacterial genus. Relative abundance was calculated only using samples containing detectable amounts of a given OTU. In instances where the OTU is only found in a single termite genus, the termite genus is also included in the label.

2.3.4 Archaeal profiles

Archaea comprise a minority of the higher termite gut community profiles with the exception of the *Macrognathotermes* sample (20% of reads) and represent >10% of the profiles in four of the nine lower termite genera investigated, in one instance comprising more than half the reads (*Porotermes* 57%; **Table 2.1**). Three of the five termite genera with high archaeal signal had multiple representatives (*Coptotermes*, *Schedorhinotermes*, and *Mastotermes*), which showed a high degree of variation in the percentage of archaeal reads (**Figure 2.3**). To cross-check that this variation and that the unexpectedly high archaeal abundance in many of these samples was not the result of primer bias, we generated additional community profiles using a different forward primer, 803F, which broadly targets bacteria and archaea (Lane, 1991). The profiles were largely consistent between the two primer sets confirming both the sample-to-sample variation within a termite genus and that the archaea comprise a high percentage of the amplicon reads in some samples (**Table 2.1** and Appendix A: **Figure S2.9**). The majority of detected archaeal phylotypes are Euryarchaeota most closely related to methanogenic genera including (in descending relative abundance) *Methanobrevibacter*, *Methanomassiliicoccus*, *Methanobacterium*, *Methanimicrococcus*, and *Methanospirillum*. Additionally, a Crenarchaeote belonging to an uncultured lineage, pGrfC26 (Großkopf et al., 1998), was detected up to 10.2% in some termite genera (**Figure 2.1** and **Figure 2.3**).

2.3.5 Eukaryotic profiles

Non-termite host eukaryotic sequences represented only 1.0% of the community profiles averaged over the 16 termite genera with the highest fraction recovered in the lower termite *Reticulitermes* (8%; **Table 2.1**). These percentages likely do not reflect protist cell numbers or ratios due to the much higher number of rRNA copies in protists relative to bacteria and variation of copy number between protist lineages (Gong et al., 2013). The majority of the eukaryotic reads were classified as parabasalids (*Trichonympha*, *Pseudotriconympha*, and *Metadevescovina*) and oxymonads (*Dinenympha*) (**Figure 2.1** and Appendix A: **Figure S2.10**). A low abundance phylotype (0.2 to 0.5%) most closely related to the ciliate *Clevelandella* (98% sequence identity) was unexpectedly detected in half of the *Gnathamitermes* samples (Appendix A: **Figure S2.10**).

2.3.6 Beta-diversity analyses

To explore the relative effect of vertical inheritance and diet on termite gut microbiota, we calculated phylogenetic distances between bacterial communities with (weighted) or without (unweighted) taking OTU relative abundance into account. Hierarchical clustering of unweighted Soergel dissimilarity distances produced a topology largely consistent with the inferred host evolution (Appendix A: **Figure S2.1**; (Legendre et al., 2008)) but not with inferred diet where dietary variation was present, that is, in polyphagous genera (**Figure 2.4** and Appendix A: **Table S2.4**). All termite genera with >1 representative were resolved as monophyletic groups according to comparison of their gut bacteria with the exception of *Coptotermes* and *Amitermes* (**Figure 2.4**). However, the latter genus was also not monophyletic within the COII tree (Appendix A: **Figure S2.1**), with FC04 and TV01 forming a separate line of descent in both trees. The vertical inheritance signal was strong enough to resolve some family level associations (with >1 genus), including the Termitidae with the exception of *Macrognathotermes* and the Kalotermitidae (**Figure 2.4**). When OTU relative abundance was taken into account, the host signal was weakened particularly at the family level, but most termite genera were still resolved as monophyletic groups (Appendix A: **Figure S2.11**). Closer inspection of *Nasutitermes* and *Gnathamitermes* revealed that relative abundance clustered members of these polyphagous genera by diet (**Figure 2.5**) reflecting the phylum-level shifts noted previously (Appendix A: **Figure S2.7**). Isotopic analysis of gut contents supports this observation as putative wood feeders had isotopically heavier carbon (d13C: -27‰ to -28‰) than their grass (d13C: -13‰ to -22‰) or dung feeding (d13C: -16‰ to -22‰) counterparts (**Figure 2.5**).

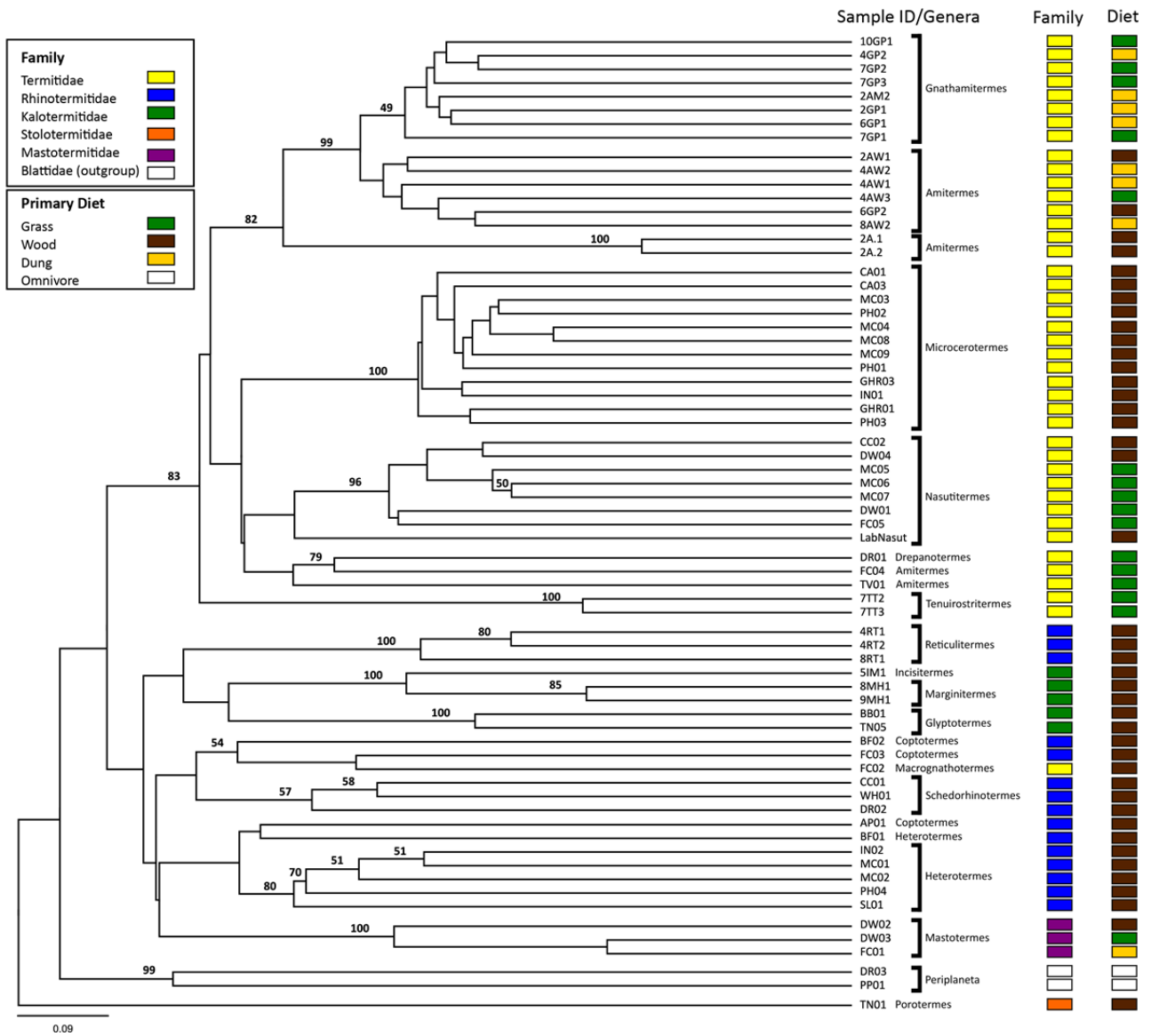


Figure 2.4: UPGMA tree of unweighted (presence/absence only) Soergel pairwise distances between bacterial profiles showing a high consistency with host phylogeny and low consistency with diet (Appendix A: Table S2.4). The values on interior nodes represent jackknife support values ≥ 49 . Termite host affiliation (family) and presumptive diet are indicated to the right of the tree.

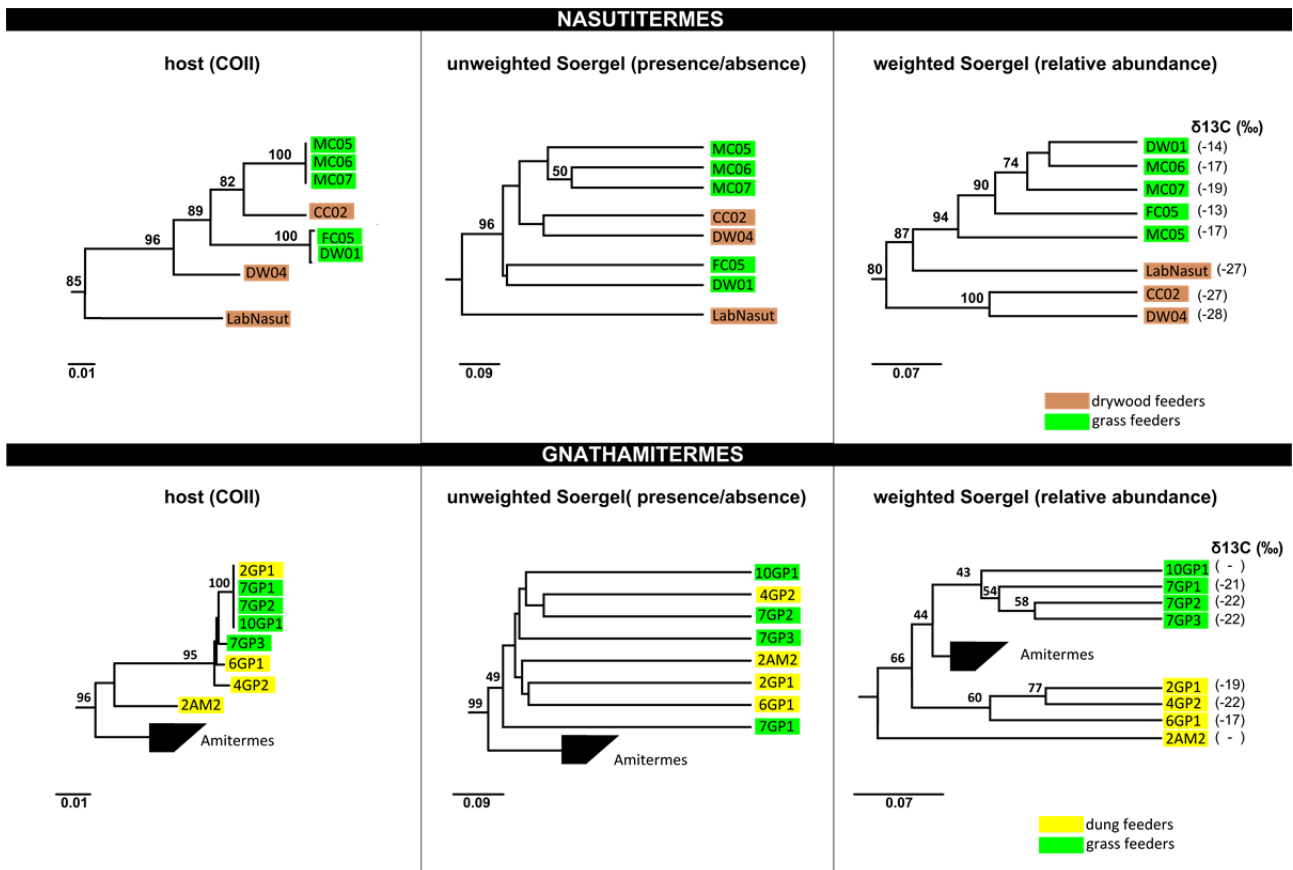


Figure 2.5: Subtrees of host and bacterial community phylogenetic comparisons showing secondary effect of diet on community structure of polyphagous termite genera. When the relative abundance of bacterial OTUs is taken into account (weighted Soergel), samples cluster according to diet. The values on interior nodes of the COII trees are FastTree local support values and jackknife support values ≥ 49 on the Soergel UPGMA trees. Carbon isotope values of gut contents are shown in the far right panels.

2.4 Discussion

Termite gut microbiota have been the subject of an increasing number of investigations over the past years using a suite of new molecular tools (Brune, 2014), however, a large amount of termite diversity remains to be explored. Here, we present the first extensive culture-independent molecular survey of the gut microbiomes of Australian termites and expand our existing knowledge of North American termite gut microbial diversity. These data are then used to assess the relative effect of vertical inheritance and environmental factors (primarily diet). The 16 termite genera examined in this study have a set of core and accessory gut bacterial phyla that distinguish them from all other habitats (**Table 2.2**). This observation is consistent with previous culture-independent studies which show that the combination of these phyla is highly distinctive of the termite gut microbiome (Warnecke et al., 2007; Hongoh, 2010; He et al., 2013; Brune, 2014; Tai et al., 2015) particularly in comparison to other insect gut communities (Douglas, 2009; Engel and Moran, 2013). This distinctiveness is further underlined by the observation that the majority of operational taxonomic units (OTUs) identified in the present study cluster with sequences from previous termite surveys (Tholen and Brune, 1999; Hongoh et al., 2003; Hongoh et al., 2005; Shinzato et al., 2005; Yang et al., 2005; Nakajima et al., 2006; Warnecke and Hugenholz, 2007; Köhler et al., 2008; Husseneder, 2010; Ikeda-Ohtsubo et al., 2010; Schauer et al., 2012). A recent extensive rRNA-based survey of gut bacteria in 34 termite species (Dietrich et al., 2014) allows direct comparison of the bacterial profiles of seven termite genera that overlap between the studies. The three higher termite profiles generally match well, but the four lower termite profiles have some conspicuous differences even at the relatively coarse phylogenetic resolution of phylum. In particular, the Dietrich et al. (2014) profiles have higher proportions of Spirochaetes and lower proportions of Bacteroidetes and Firmicutes than the corresponding profiles in our study (Appendix A: **Figure S2.12**). For *Reticulitermes* and *Coptotermes*, this may reflect real differences as different species were profiled, but for *Mastotermes* and *Incisitermes* for which the same species were examined, the more likely explanation is differences arising from methodology such as DNA extraction method (Morgan et al., 2010) and/or PCR primers used (Engelbrektson et al., 2010). A study by Sabree and Moran (Sabree and Moran, 2014) using similar DNA extraction method and primers to ours produced a similar gut community profile for *Mastotermes* (Appendix A: **Figure S2.12**).

With these methodological caveats in mind, key differences between higher and lower termite gut profiles are linked to the presence of protists in the latter group. The Bacteroidetes and Elusimicrobia are the most over-represented phyla in lower termites relative to higher termites because they harbor highly abundant members of the *Candidatus* genera Azobacteroides,

Vestibaculum (Bacteroidetes), and Endomicrobium (Elusimicrobia; **Figure 2.2**), which are recognised protist symbionts (Noda et al., 2005; Noda et al., 2006; Noda et al., 2007; Stingl et al., 2004; Stingl et al., 2005). *Candidatus Azobacteroides pseudotrichonymphae*, an endosymbiont of the parabasalid *Pseudotrichonympha grasioi*, has previously been reported to comprise approximately 70% of the bacterial cells present in the gut of *Coptotermes formosanus* (Hongoh et al., 2008). Here, we found phylogenetically distinct *Candidatus Azobacteroides* spp. comprise up to 66% of the bacterial gut profiles in *Coptotermes*, up to 63% in *Schedorhinotermes*, and up to 72% in *Heterotermes* (Appendix A: **Figure S2.8**) and identified their putative *Pseudotrichonympha* hosts only in those termite genera (**Figure 2.1** and Appendix A: **Figure S2.10**), supporting the previously reported specific relationship between the two in multiple termite genera (Noda et al., 2007). *Candidatus Vestibaculum illigatum* was first reported in *Neotermes cubanus* and was shown to be an epibiont of the flagellated protist *Staurojoenina* (Stingl et al., 2004). Here, we found abundant populations (8%–22%) of *Candidatus Vestibaculum* in *Incisitermes* and *Marginitermes*, both members of the family Kalotermitidae. Assuming that *Candidatus Vestibaculum* is a specific epibiont of *Staurojoenina*, this is consistent with the observation that *Staurojoenina* is only found in members of the family Kalotermitidae (Yamin, 1979; Gile et al., 2013). However, the other Kalotermitidae genus surveyed, *Glyptotermes*, lacked detectable populations of *Candidatus Vestibaculum* (Appendix A: **Figure S2.8**) and *Staurojoenina* was not detected at all in our survey. The latter observation may be due to our primer set not targeting this parabasalid genus (two mismatches in the 926F primer to *S. assimilis* acc. AB183882).

Candidatus Endomicrobium was detected in all of the lower termite genera surveyed and was also found in low abundance in some of the higher termite genera (**Figure 2.1**) consistent with previous findings (Stingl et al., 2005; Ikeda-Ohtsubo et al., 2007; Ohkuma et al., 2007). In *Reticulitermes* and *Incisitermes*, *Candidatus Endomicrobium* is a recognised cytoplasmic symbiont of the parabasalids *Trichonympha* and *Metadevescovina*, respectively (Ohkuma et al., 2007). Our data are consistent with these observations as high abundance populations of *Candidatus Endomicrobium*, and their respective host protists were detected in *Reticulitermes* and *Incisitermes* (**Figure 2.1**). The highest relative abundance of *Candidatus Endomicrobium* was found in *Porotermes* (65% of bacterial reads; **Figure 2.2** and Appendix A: **Figure S2.8**), however, no protist host sequences were detected, presumably due to primer mismatches as visual observation of *Porotermes* gut contents reveal a high diversity of protist morphotypes (unpublished observation).

The most prevalent (ubiquitous) genus in the gut survey was *Treponema* (**Figure 2.2**), which comprises most of the Spirochaetes phylum signal. *Treponema* has been reported in every termite gut investigation to date (Brune, 2014) and contributes substantially to the distinctiveness of the

termite gut microbiome. Numerous *Treponema* OTUs were found in the present survey, many of which flourished in the higher termites (Appendix A: **Figure S2.8**) likely following the evolutionary loss of protists from the hindgut (Brandl et al., 2007). It has been shown that Spirochaetes are essential for the survival of higher termites and that their removal results in a shorter life span (Eutick et al., 1978). Metagenomic, metatranscriptomic, and metaproteomic analyses of two higher termite genera, *Nasutitermes* and *Amitermes*, indicate that treponemes are involved in all of the major functions in the hindgut, including fibre hydrolysis, fermentation, homoacetogenesis, and nitrogen fixation (Warnecke et al., 2007; He et al., 2013) which may explain their success (ubiquity) and long term co-habitation with their termite hosts. However, *Treponema* is a phylogenetically broad genus (Ohkuma et al., 1999a; Breznak and Leadbetter, 2006; Berlanga et al., 2009) and it seems likely that not all species will be capable of all key functions.

Two deltaproteobacteria (*Desulfovibrio* and *Desulfarculus*-like OTUs) were present in low abundance in over half of the samples tested (**Figure 2.2**). *Desulfovibrio* has previously been reported as a widespread constituent of termite guts mainly based on cultivation studies, with proposed functions including oxygen removal and nitrogen fixation (Kuhnigk et al., 1996). However, the *Desulfarculus*-like OTU was more prevalent (Appendix A: **Figure S2.8**) and a member of this group has recently been inferred to be primarily responsible for the first step in CO₂-reductive acetogenesis (Rosenthal et al., 2013). If this key functionality in the *Desulfarculus*-like group is conserved across different termite genera, it may explain their widespread distribution among the surveyed termites. Less expected was the widespread occurrence of a *Pirellula*-like planctomycete OTU (**Figure 2.2**). Planctomycetes have been reported in alkaline gut segments of soil-feeding termite genera, where they are speculated to play a role in degradation of humus-associated biopolymers such as N-acetylglucosamine (Köhler et al., 2008). No soil-feeding genera were surveyed in the present study, although the planctomycete OTU may be associated with alkaline segments known to be present in several higher termite genera (Bignell and Eggleton, 1995). The planctomycete OTU was also detected in three lower termite genera (Appendix A: **Figure S2.8**) which are not known to have alkaline gut segments, suggesting that planctomycetes are not strictly associated with higher pH in termites (Hongoh et al., 2003; Hongoh et al., 2005; Nakajima et al., 2005).

Archaea have been reported to constitute only a small fraction (up to 3%) of the termite gut ecosystem (Brauman et al., 2001), however, we found much higher percentages in the amplicon profiles of a number of lower termite genera and one higher termite genus (**Table 2.1**). We cross-checked our findings with an alternative forward primer broadly targeting bacteria and archaea (803F) and confirmed that the result was not an artefact of the universal primer pair (926F and

1392R). Also considering that many samples had archaeal proportions in the anticipated range (<3%; **Table 2.1**), we suggest that the higher values are not artefacts of the primers or of the DNA extraction method used. The observed variability in archaeal abundance between samples belonging to the same termite genus, e.g. *Schedorhinotermes* (1.4%, 24.5%, and 32.9%), suggests that archaeal abundance may be more variable between specimens than previously appreciated, possibly reflecting environmental factors or simply temporal dynamics ('archaeal blooms'). Only hydrogenotrophic methanogens, dominated by *Methanobrevibacter* in most cases, were detected in the surveyed termite guts consistent with previous reports (Ohkuma et al., 1999b; Tokura et al., 2000; Shinzato et al., 2001), suggesting that acetoclastic methanogenesis is likely unfavourable in this habitat. Phylotypes closely related to a recently described phylogenetically novel methanogenic genus related to the Thermoplasmatales, *Methanomassiliicoccus*, were detected in several termite genera raising the possibility that these methanogenic populations may have an obligate requirement for methanol (Paul et al., 2012).

Eukaryotes were not the primary focus of this study, and our data are likely an underestimate of protist diversity in the surveyed species due to primer mismatches (Hadziavdic et al., 2014; Wang et al., 2014). Also, rRNA-based relative abundance estimates will likely not reflect cell counts (e.g. *Reticulitermes* (Lewis and Forschler, 2004)) due to the much higher number of rRNA gene copies in protists relative to bacteria (Gong et al., 2013), the former of which is not currently corrected by CopyRighter (Angly et al., 2014). However, some interesting qualitative observations were made including putative protist host-bacterial symbiont pairings described above. It is commonly reported that higher termites lack flagellated protists, which are primary agents of lignocellulose digestion in lower termites (Brune and Ohkuma, 2011; Hongoh, 2011; Brune, 2014). Instead, bacteria and to a lesser extent, the termite itself, provide the enzymes necessary for lignocellulose hydrolysis in higher termites (Brune, 2014; Hongoh, 2014). Unexpectedly then, a phylotype related to the ciliate *Clevelandella*, previously reported in wood-feeding cockroach intestinal tracts (Lynn and Wright, 2013), was detected in the higher termite genus *Gnathamitermes* (Appendix A: **Figure S2.10**). An older microscopic study of higher termite gut ecosystems supports our findings with the identification of small numbers of a closely related ciliate, *Nyctotherus*, in *Amitermes* (Kirby, 1932), although no protists were detected in the *Amitermes* community profiles in the present study. The presence of low abundance protist populations in some higher termite genera suggests either reacquisition after the major evolutionary transition to bacteria-dominated gut communities in the higher termites or low-level persistence of some protist species across this transition. It will be interesting to determine if these ciliates are directly involved in lignocellulose digestion.

A primary motivation of our study was to determine the relative importance of vertical inheritance (host signal) versus diet on termite gut microbiota composition given the unusual status of termites among insects in terms of gut microbiome complexity (Douglas, 2011) and the importance of termites as ecosystem engineers (He et al., 2013). This question is not immediately addressable using field observations of lower termites as they are primarily wood feeders with the exception of *Mastotermes* (Abe et al., 2000; Andersen and Jacklyn, 1993). However, we obtained sufficient specimens of polyphagous higher termite genera to evaluate the relative effect of diet and host signal. The strongest signal was clearly due to vertical inheritance, with termite genus and even family level associations being resolved based on gut community profiles alone, particularly in unweighted analyses (**Figure 2.4**). This is consistent with previous studies indicating that vertical transmission plays an important role in structuring termite gut communities, for example co-speciation of gut symbionts within the genera *Reticulitermes* and *Microcerotermes* (Hongoh et al., 2005) and a general host signal in whole gut community analyses of 34 termite and cockroach species (Dietrich et al., 2014). Maintenance of host-specific microbial communities must be achieved via vertical transmission during trophallaxis or coprophagy, as there is no germline transfer in termites (Huang et al., 2008). It is important to note that a dominant host signal in gut community composition does not imply that all component species are the product of vertical inheritance, ultimately resulting in co-speciation. The termite gut is an open system that would allow ingress of foreign microorganisms, which may be able to persist under favourable conditions. For example, it was speculated that some Firmicute populations in *Amitermes* have been laterally acquired from herbivore gut communities as a result of dietary specialisation, i.e. dung feeding (He et al., 2013) (*see below*). These bacterial populations were then subsequently vertically transmitted in the *Amitermes* lineage. While fine-scale reconstruction of population co-evolution is not feasible with partial rRNA sequences, the clusters of identical reads identified in the most ubiquitous 97% OTU, *Treponema* OTU1, reflects the dominant overall host signal but also suggests that a minority of strains in the cluster may have been laterally transferred between termite genera (Appendix A: **Table S2.3**).

The effect of diet on gut community structure has been addressed to a lesser extent in termites. No clear dietary signal was observed in unweighted analyses (**Figure 2.4**), but when the evenness (relative abundance) of gut phylotypes was taken into account, a secondary effect of diet on community structure became apparent in the well-sampled polyphagous termite genera. Specifically, *Nasutitermes* samples are partitioned into wood- and grass-feeding clades and *Gnathamitermes* into grass- and dung-feeding clades (**Figure 2.5**). Changes in phylum-level abundances could be correlated with the dietary differences such as an increased abundance of

Spirochaetes and Fibrobacteres and decreased abundance of Firmicutes in wood-feeding relative to grass-feeding *Nasutitermes* (Appendix A: **Figure S2.7**). This is consistent with previous reports of the importance of Spirochaetes and Fibrobacteres in the digestion of wood fibres (Warnecke et al., 2007; Mikaelyan et al., 2014). He et al. (He et al., 2013) identified phylum-level shifts between dung-feeding *Amitermes* and wood-feeding *Nasutitermes*. Based on metagenomic and metatranscriptomic analyses, they explained these differences by inferring that Firmicutes play a greater role in hemicellulose hydrolysis and utilisation of fixed-nitrogen compounds required for dung digestion and Spirochaetes play a greater role in cellulose hydrolysis and nitrogen fixation required for wood digestion. However, our data suggest that phylum-level differences attributed to diet were overestimated in the He et al. study (He et al., 2013) because of marked differences between the *Amitermes* and *Nasutitermes* gut communities due to vertical inheritance. We estimate that changes in the relative abundance of these phyla between wood- and grass-feeding *Nasutitermes* samples is only 4%–8%, as opposed to the 15%–34% differences seen between dung-feeding *Amitermes* and wood-feeding *Nasutitermes* (Appendix A: **Figure S2.7**). Presumably in some instances, changes in relative abundances of gut populations occurred over evolutionary time scales in response to dedicated dietary specialisation (Andersen and Jacklyn, 1993). However, recent feeding trial studies of *Reticulitermes flavipes* indicate that such changes in population evenness can occur on short time scales allowing polyphagous termite species to adjust rapidly to changes in their diet due to seasonal variation or availability of foraged plant species (Boucias et al., 2013; Huang et al., 2013).

2.5 Conclusions

In summary, we infer that vertical inheritance is the primary force shaping the termite gut microbiome and that most indigenous species are successfully and faithfully passed from one termite generation to the next. Changes in relative abundance can occur on shorter time scales and appear to be an adaptive mechanism for changes in diet. The resilience of termite gut communities to experimental dietary perturbations remains to be fully explored. Our findings suggest that an evolutionary perspective will greatly assist in deconvoluting specific and whole community functionality in the termite gut microbiome.

2.6 Acknowledgements

We thank the following colleagues for assistance with collection and processing of specimens used in this study: Martin Allgaier, Gary Cochrane, John Gosper, Patrick Keeling, Scott Kleinschmidt, Victor Kunin, Linda Ly, Lisa Margonelli, Mark Morrison, Micheal Neal, Peter O'Donoghue, Carly Rosewarne, Rochelle Soo, and Brian Thistleton. We thank Shaomei He and Susannah Tringe for sample sorting and shipping of US termite samples to Australia; Fiona May, Sue Read, and members of the JGI production team for assistance with pyrosequencing; Stephanie Malfatti for assistance with sequence analysis; Lyn Cook for useful discussions on marker genes; and Ray Lee for isotope analysis. The study was supported by funding to the Australian Centre for Ecogenomics and the DOE Joint Genome Institute and Queensland Smart Futures Fund (Future biofuels). NR was supported by a UQ Research Scholarship. We dedicate this manuscript to the memory of our friend and colleague Falk Warnecke (1972–2014).

2.7 References

- Abe, T., Bignell, D.E., and Higashi, M. (2000) *Termites: Evolution, Sociality, Symbioses, Ecology*: Kluwer Academic Publishers.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.
- Andersen, A.N., and Jacklyn, P. (1993) *Termites of the top end*: CSIRO Publishing.
- Angly, F.E., Dennis, P.G., Skarshewski, A., Vanwonderghem, I., Hugenholtz, P., and Tyson, G.W. (2014) CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome* **2**: 11.
- Bandi, C., Anderson, T.J., Genchi, C., and Blaxter, M.L. (1998) Phylogeny of *Wolbachia* in filarial nematodes. *Proceedings of the Royal Society of London Series B: Biological Sciences* **265**: 2407-2413.
- Berlanga, M., Paster, B.J., and Guerrero, R. (2009) The taxophysiological paradox: changes in the intestinal microbiota of the xylophagous cockroach *Cryptocercus punctulatus* depending on the physiological state of the host. *International Microbiology* **12**: 227-236.
- Bignell, D.E., and Eggleton, P. (1995) On the elevated intestinal pH of higher termites (Isoptera: Termitidae). *Insectes Sociaux* **42**: 57-69.
- Boucias, D.G., Cai, Y., Sun, Y., Lietze, V.U., Sen, R., Raychoudhury, R., and Scharf, M.E. (2013) The hindgut lumen prokaryotic microbiota of the termite *Reticulitermes flavipes* and its responses to dietary lignocellulose composition. *Molecular Ecology* **22**: 1836-1853.
- Bragg, L., Stone, G., Imelfort, M., Hugenholtz, P., and Tyson, G.W. (2012) Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nature Methods* **9**: 425-426.

- Brandl, R., Hyodo, F., Korff-Schmising, M.v., Maekawa, K., Miura, T., Takematsu, Y. et al. (2007) Divergence times in the termite genus *Macrotermes* (Isoptera: Termitidae). *Molecular Phylogenetics and Evolution* **45**: 239-250.
- Brauman, A., Dore, J., Eggleton, P., Bignell, D., Breznak, J.A., and Kane, M.D. (2001) Molecular phylogenetic profiling of prokaryotic communities in guts of termites with different feeding habits. *FEMS Microbiology Ecology* **35**: 27-36.
- Breznak, J.A., and Leadbetter, J.R. (2006) Termite gut spirochetes. In *The prokaryotes*: Springer, pp. 318-329.
- Brune, A. (2014) Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology* **12**: 168-180.
- Brune, A., and Ohkuma, M. (2011) Role of the Termite Gut Microbiota in Symbiotic Digestion. In *Biology of Termites: a Modern Synthesis*. Bignell, D.E., Roisin, Y., and Lo, N. (eds): Springer Netherlands, pp. 439-475.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335-336.
- Dietrich, C., Köhler, T., and Brune, A. (2014) The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Applied and Environmental Microbiology* **80**: 2261-2269.
- Douglas, A. (2009) The microbial dimension in insect nutritional ecology. *Functional Ecology* **23**: 38-47.
- Douglas, A.E. (2011) Lessons from studying insect symbioses. *Cell Host Microbe* **10**: 359-367.
- Engel, P., and Moran, N.A. (2013) The gut microbiota of insects—diversity in structure and function. *FEMS Microbiology Reviews* **37**: 699-735.
- Engelbrekton, A., Kunin, V., Wrighton, K.C., Zvenigorodsky, N., Chen, F., Ochman, H., and Hugenholtz, P. (2010) Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *International Society for Microbial Ecology Journal* **4**: 642-647.
- Eutick, M.L., Veivers, P., O'Brien, R.W., and Slaytor, M. (1978) Dependence of the higher termite, *Nasutitermes exitiosus* and the lower termite, *Coptotermes lacteus* on their gut flora. *Journal of Insect Physiology* **24**: 363-368.
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**: 3150-3152.
- Gile, G.H., Carpenter, K.J., James, E.R., Scheffrahn, R.H., and Keeling, P.J. (2013) Morphology and molecular phylogeny of *Staurojoenina mulleri* sp. nov. (Trichonymphida, Parabasalia) from the hindgut of the Kalotermitid *Neotermes jouteli*. *Journal of Eukaryotic Microbiology* **60**: 203-213.
- Gong, J., Dong, J., Liu, X., and Massana, R. (2013) Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and peritrich ciliates. *Protist* **164**: 369-379.

- Großkopf, R., Stubner, S., and Liesack, W. (1998) Novel euryarchaeotal lineages detected on rice roots and in the anoxic bulk soil of flooded rice microcosms. *Applied and Environmental Microbiology* **64**: 4983-4989.
- Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M., and Troedsson, C. (2014) Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PloS One* **9**: e87624.
- He, S., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R.H. et al. (2013) Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS One* **8**: e61126.
- Hongoh, Y. (2010) Diversity and genomes of uncultured microbial symbionts in the termite gut. *Bioscience, Biotechnology, and Biochemistry* **74**: 1145-1151.
- Hongoh, Y. (2011) Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cellular and Molecular Life Sciences* **68**: 1311-1325.
- Hongoh, Y. (2014) Who digests the lignocellulose? *Environmental Microbiology* **9**:2644-2645.
- Hongoh, Y., Deevong, P., Inoue, T., Moriya, S., Trakulnaleamsai, S., Ohkuma, M. et al. (2005) Intra- and interspecific comparisons of bacterial diversity and community structure support coevolution of gut microbiota and termite host. *Applied and Environmental Microbiology* **71**: 6590-6599.
- Hongoh, Y., Ohkuma, M., and Kudo, T. (2003) Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera; Rhinotermitidae). *FEMS Microbiology Ecology* **44**: 231-242.
- Hongoh, Y., Sharma, V.K., Prakash, T., Noda, S., Toh, H., Taylor, T.D. et al. (2008) Genome of an endosymbiont coupling N₂ fixation to cellulolysis within protist cells in termite gut. *Science* **322**: 1108-1109.
- Huang, Q.-Y., Wang, W.-P., Mo, R.-Y., and Lei, C.-L. (2008) Studies on feeding and trophallaxis in the subterranean termite *Odontotermes formosanus* using rubidium chloride. *Entomologia Experimentalis et Applicata* **129**: 210-215.
- Huang, X.-F., Bakker, M.G., Judd, T.M., Reardon, K.F., and Vivanco, J.M. (2013) Variations in diversity and richness of gut bacterial communities of termites (*Reticulitermes flavipes*) fed with grassy and woody plant substrates. *Microbial Ecology* **65**: 531-536.
- Husseneder, C. (2010) Symbiosis in subterranean termites: a review of insights from molecular studies. *Environmental Entomology* **39**: 378-388.
- Ikeda-Ohtsubo, W., Faivre, N., and Brune, A. (2010) Putatively free-living 'Endomicrobia'-ancestors of the intracellular symbionts of termite gut flagellates? *Environmental Microbiology Reports* **2**: 554-559.
- Ikeda-Ohtsubo, W., Desai, M., Stingl, U., and Brune, A. (2007) Phylogenetic diversity of 'Endomicrobia' and their specific affiliation with termite gut flagellates. *Microbiology* **153**: 3458-3465.

- Inward, D.J., Vogler, A.P., and Eggleton, P. (2007) A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology. *Molecular Phylogenetics and Evolution* **44**: 953-967.
- Kirby, H. (1932) Protozoa in termites of the genus *Amitermes*. *Parasitology* **24**: 289-304.
- Köhler, T., Stingl, U., Meuser, K., and Brune, A. (2008) Novel lineages of *Planctomycetes* densely colonize the alkaline gut of soil-feeding termites (*Cubitermes* spp.). *Environmental Microbiology* **10**: 1260-1270.
- Krishna, K., Grimaldi, D.A., Krishna, V., and Engel, M.S. (2013) Treatise on the Isoptera of the world. *Bulletin of the American Museum of Natural History* **377**: 200-623.
- Kuhnigk, T., Branke, J., Krekeler, D., Cypionka, H., and König, H. (1996) A feasible role of sulfate-reducing bacteria in the termite gut. *Systematic and Applied Microbiology* **19**: 139-149.
- Kunin, V., and Hugenholtz, P. (2010) PyroTagger : A fast , accurate pipeline for analysis of rRNA amplicon pyrosequence data. *The Open Journal*: 1-8.
- Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* **12**: 118-123.
- Lane, D.J. (1991) Nucleic Acid Techniques in Bacterial Systematics. In *Nucleic Acid Techniques in Bacterial Systematics*. M., S.E.G. (ed): Wiley, New York, pp. 115-175.
- Legendre, F., Whiting, M.F., Bordereau, C., Canello, E.M., Evans, T.A., and Grandcolas, P. (2008) The phylogeny of termites (Dictyoptera: Isoptera) based on mitochondrial and nuclear markers: Implications for the evolution of the worker and pseudergate castes, and foraging behaviors. *Molecular Phylogenetics and Evolution* **48**: 615-627.
- Lewis, J.L., and Forschler, B.T. (2004) Protist communities from four castes and three species of *Reticulitermes* (Isoptera: Rhinotermitidae). *Annals of the Entomological Society of America* **97**: 1242-1251.
- Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R., and Gordon, J.I. (2008a) Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology* **6**: 776-788.
- Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S. et al. (2008b) Evolution of mammals and their gut microbes. *Science* **320**: 1647-1651.
- Liu, H., and Beckenbach, A.T. (1992) Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. *Molecular Phylogenetics and Evolution* **1**: 41-52.
- Lozupone, C., and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology* **71**: 8228-8235.
- Lozupone, C., Hamady, M., and Knight, R. (2006) UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**: 371.
- Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., and Knight, R. (2011) UniFrac: an effective distance metric for microbial community comparison. *The ISME Journal* **5**: 169-172.

- Lynn, D.H., and Wright, A.-D.G. (2013) Biodiversity and molecular phylogeny of Australian *Clevelandella* species (Class *Armophorea*, Order *Clevelandellida*, Family *Clevelandellidae*), intestinal endosymbiotic ciliates in the wood-feeding roach *Panesthia cribrata* Saussure, 1864. *Journal of Eukaryotic Microbiology* **60**: 335-341.
- McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A. et al. (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal* **6**: 610-618.
- Mikaelyan, A., Strassert, J.F., Tokuda, G., and Brune, A. (2014) The fibre-associated cellulolytic bacterial community in the hindgut of wood-feeding higher termites (*Nasutitermes* spp.). *Environmental Microbiology*.
- Moran, N.A., and Mira, A. (2001) The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biology* **2**: 1-0054.0012.
- Morgan, J.L., Darling, A.E., and Eisen, J.A. (2010) Metagenomic sequencing of an in vitro-simulated microbial community. *PLoS One* **5**: e10209.
- Nakajima, H., Hongoh, Y., Noda, S., Yoshida, Y., Usami, R., Kudo, T., and Ohkuma, M. (2006) Phylogenetic and morphological diversity of *Bacteroidales* members associated with the gut wall of termites. *Bioscience, Biotechnology, and Biochemistry* **70**: 211-218.
- Nakajima, H., Hongoh, Y., Usami, R., Kudo, T., and Ohkuma, M. (2005) Spatial distribution of bacterial phylotypes in the gut of the termite *Reticulitermes speratus* and the bacterial community colonizing the gut epithelium. *FEMS Microbiology Ecology* **54**: 247-255.
- Noda, S., Iida, T., Kitade, O., Nakajima, H., Kudo, T., and Ohkuma, M. (2005) Endosymbiotic *Bacteroidales* bacteria of the flagellated protist *Pseudotriconympha grassii* in the gut of the termite *Coptotermes formosanus*. *Applied and Environmental Microbiology* **71**: 8811-8817.
- Noda, S., Inoue, T., Hongoh, Y., Kawai, M., Nalepa, C.A., Vongkaluang, C. et al. (2006) Identification and characterization of ectosymbionts of distinct lineages in *Bacteroidales* attached to flagellated protists in the gut of termites and a wood-feeding cockroach. *Environmental Microbiology* **8**: 11-20.
- Noda, S., Kitade, O., Inoue, T., Kawai, M., Kanuka, M., Hiroshima, K. et al. (2007) Cospeciation in the triplex symbiosis of termite gut protists (*Pseudotriconympha* spp.), their hosts, and their bacterial endosymbionts. *Molecular Ecology* **16**: 1257-1266.
- Ochman, H., Worobey, M., Kuo, C.H., Ndjango, J.B., Peeters, M., Hahn, B.H., and Hugenholtz, P. (2010) Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biology* **8**: e1000546.
- Ohkuma, M., Iida, T., and Kudo, T. (1999a) Phylogenetic relationships of symbiotic spirochetes in the gut of diverse termites. *FEMS Microbiology Letters* **181**: 123-129.
- Ohkuma, M., Noda, S., and Kudo, T. (1999b) Phylogenetic diversity of nitrogen fixation genes in the symbiotic microbial community in the gut of diverse termites. *Applied and Environmental Microbiology* **65**: 4926-4934.

- Ohkuma, M., Sato, T., Noda, S., Ui, S., Kudo, T., and Hongoh, Y. (2007) The candidate phylum 'Termite Group 1' of bacteria: phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiology Ecology* **60**: 467-476.
- Ohkuma, M., Yuzawa, H., Amornsak, W., Sornnuwat, Y., Takematsu, Y., Yamada, A. et al. (2004) Molecular phylogeny of Asian termites (Isoptera) of the families Termitidae and Rhinotermitidae based on mitochondrial COII sequences. *Molecular Phylogenetics and Evolution* **31**: 701-710.
- Parks, D.H., and Beiko, R.G. (2010) Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* **26**: 715-721.
- Parks, D.H., and Beiko, R.G. (2013) Measures of phylogenetic differentiation provide robust and complementary insights into microbial communities. *The ISME Journal* **7**: 173-183.
- Patin, N., Kunin, V., Lidström, U., and Ashby, M. (2013) Effects of OTU clustering and PCR artifacts on microbial diversity estimates. *Microbial Ecology* **65**: 709-719.
- Paul, K., Nonoh, J.O., Mikulski, L., and Brune, A. (2012) "Methanoplasmatales," *Thermoplasmatales*-related archaea in termite guts and other environments, are the seventh order of methanogens. *Applied and Environmental Microbiology* **78**: 8245-8253.
- Price, M.N., Dehal, P.S., and Arkin, A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution* **26**: 1641-1650.
- Rosenthal, A.Z., Zhang, X., Lucey, K.S., Ottesen, E.A., Trivedi, V., Choi, H.M. et al. (2013) Localizing transcripts to single cells suggests an important role of uncultured deltaproteobacteria in the termite gut hydrogen economy. *Proceedings of the National Academy of Sciences* **110**: 16163-16168.
- Sabree, Z.L., and Moran, N.A. (2014) Host-specific assemblages typify gut microbial communities of related insect species. *SpringerPlus* **3**: 138.
- Schauer, C., Thompson, C.L., and Brune, A. (2012) The bacterial community in the gut of the cockroach *Shelfordella lateralis* reflects the close evolutionary relatedness of cockroaches and termites. *Applied and Environmental Microbiology* **78**: 2758-2767.
- Scheffrahn, R.H., and Postle, A. (2013) New termite species and newly recorded genus for Australia: *Marginitermes absitus* (Isoptera: Kalotermitidae). *Australian Journal of Entomology* **52**: 199-205.
- Shinzato, N., Muramatsu, M., Matsui, T., and Watanabe, Y. (2005) Molecular phylogenetic diversity of the bacterial community in the gut of the termite *Coptotermes formosanus*. *Bioscience, Biotechnology, and Biochemistry* **69**: 1145-1155.
- Shinzato, N., Matsumoto, T., Yamaoka, I., Oshima, T., and Yamagishi, A. (2001) Methanogenic symbionts and the locality of their host lower termites. *Microbes and Environments* **16**: 43-47.
- Stingl, U., Maass, A., Radek, R., and Brune, A. (2004) Symbionts of the gut flagellate *Staurojoenina* sp. from *Neotermes cubanus* represent a novel, termite-associated lineage of *Bacteroidales*: description of 'Candidatus Vestibaculum illigatum'. *Microbiology* **150**: 2229-2235.
- Stingl, U., Radek, R., Yang, H., and Brune, A. (2005) "Endomicrobia": cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Applied and Environmental Microbiology* **71**: 1473-1479.

- Tai, V., James, E.R., Nalepa, C.A., Scheffrahn, R.H., Perlman, S.J., and Keeling, P.J. (2015) The role of host phylogeny varies in shaping microbial diversity in the hindguts of lower termites. *Applied and Environmental Microbiology* **81**: 1059-1070.
- Tholen, A., and Brune, A. (1999) Localization and in situ activities of homoacetogenic bacteria in the highly compartmentalized hindgut of soil-feeding higher termites (*Cubitermes* spp.). *Applied and Environmental Microbiology* **65**: 4497-4505.
- Thompson, G., Kitade, O., Lo, N., and Crozier, R. (2000) Phylogenetic evidence for a single, ancestral origin of a 'true' worker caste in termites. *Journal of Evolutionary Biology* **13**: 869-881.
- Tokura, M., Ohkuma, M., and Kudo, T. (2000) Molecular phylogeny of methanogens associated with flagellated protists in the gut and with the gut epithelium of termites. *FEMS Microbiology Ecology* **33**: 233-240.
- Turnbaugh, P.J., Quince, C., Faith, J.J., McHardy, A.C., Yatsunenko, T., Niazi, F. et al. (2010) Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proceedings of the National Academy of Sciences* **107**: 7503-7508.
- Wang, Y., Tian, R.M., Gao, Z.M., Bougouffa, S., and Qian, P.-Y. (2014) Optimal eukaryotic 18S and universal 16S/18S ribosomal RNA primers and their application in a study of symbiosis. *PLoS One* **9**: e90053.
- Ware, J.L., Grimaldi, D.A., and Engel, M.S. (2010) The effects of fossil placement and calibration on divergence times and rates: an example from the termites (Insecta: Isoptera). *Arthropod Structure & Development* **39**: 204-219.
- Warnecke, F., and Hugenholtz, P. (2007) Building on basic metagenomics with complementary technologies. *Genome Biology* **8**: 231.
- Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T. et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**: 560-565.
- Yamin MA (1979) Flagellates of the orders Trichomonadida Kirby, Oxymonadida Grasse, and Hypermastigida Grassi & Foa reported from lower termites (Isoptera families Mastotermitidae, Kalotermitidae, Hodotermitidae, Termopsidae, Rhinotermitidae, and Serritermitidae) and from the wood-feeding roach *Cryptocercus* (Dictyoptera: Cryptocercidae). *Sociobiology* **4**: 5-119
- Yang, H., Schmitt-Wagner, D., Stingl, U., and Brune, A. (2005) Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environmental Microbiology* **7**: 916-932.

Chapter 3 Effect of diet on microbial community composition in the gut of *Mastotermes darwiniensis*

Abstract

From our initial amplicon survey, we hypothesised that dietary fluctuations caused shifts in the gut community structure and function. We directly tested this hypothesis through a series of feeding experiments performed on the most basal polyphagous termite species *Mastotermes darwiniensis*. Here, using amplicon community profiling and proteome analyses, we investigated the *M. darwiniensis* gut microbial community compositional changes in response to wood (complex lignocellulose), sugarcane mulch (C4 grass) and cotton (pure cellulose) diet over a period of seven days. Shifts in relative abundance of some gut microbial populations were noted with compositionally different feedstocks but most of the changes are likely due to response to stress as an effect of small colonies. While laboratory conditions can influence the termite gut microbiome, the bulk of prokaryotic communities of *Mastotermes* gut profiles were comparably consistent in all diets. In addition, zymography coupled with mass spectrometry sequencing of *M. darwiniensis* gut extracts identified highly expressed protist-derived cellulases. Thus, our findings have contributed to the understanding of community structure and function of gut microbiota in lower termites and suggest that the contribution of prokaryotic symbionts to lignocellulose digestion is more complex than previously appreciated.

3.1 Introduction

Microbial communities are generally sensitive to disturbances, despite the system or the type of disturbance, and change in composition may affect the overall dynamics of an ecosystem (Allison and Martiny, 2008). Community stability is often defined as the ability of the community to maintain its functional and structural properties in response to perturbation (Ives and Carpenter, 2007; Little et al., 2008; Robinson et al., 2010). Ecological concepts of community stability have been applied when evaluating how perturbation can affect microbial composition (Robinson et al., 2010). Three factors are considered when determining the extent to which microbial communities respond to environmental perturbations (1) *resistance*; the capacity of a microbial community to withstand perturbations without loss of structure or function, (2) *resilience*; the ability of a community to return to its original composition after disturbance, and (3) *functional redundancy*; the ability of a community to perform the same functions despite compositional changes after disturbance (Allison and Martiny, 2008).

Gastrointestinal tract microbial communities are influenced by various factors (perturbations) including changes in diet. A number of studies have found that both host phylogeny and diet play a role in shaping gut microbial communities as observed in mammals (Ley et al., 2008; Muegge et al., 2011; Wu et al., 2011), arthropods (Tanaka et al., 2006; Husseneder et al., 2009; Cardoso et al., 2012) and fish (Ward et al., 2009; Smriga et al., 2010; Miyake et al., 2015). Our findings in Chapter 2 indicate that in the termite, diet plays a secondary role in shaping community composition after host phylogeny, and we predicted that diet may alter relative abundances of populations (evenness) on shorter timescales as an adaptive mechanism for varying diets in polyphagous termite genera such as *Nasutitermes*, *Gnathamitermes* and *Mastotermes*. Due to their ability to feed on diverse complex lignocellulose biomass as reported in Chapter 2, polyphagous termite species serve as an excellent model system to test host-associated gut community stability in response to changes in diet.

Mastotermes darwiniensis is the only recognised polyphagous lower termite species that is indigenous to Australia. Different colonies of *M. darwiniensis* are known to feed on different types of biomass ranging from drywood, palm trees, cattle dung to soil (Andersen and Jacklyn, 1993), yet little is known about the response of the gut community to changes in diet. Although recent feeding experiments in lower termites suggest that gut microbial communities are sensitive to diet, these assays were conducted on strict wood-feeding termite species and only profiled gut community composition at the end of a feeding trial (Husseneder et al., 2009; Sethi et al., 2012; Boucias et al., 2013; Huang et al., 2013; Raychoudhury et al., 2013). Here, using culture-independent SSU rRNA

amplicon-based sequencing, we performed a perturbation study to test the hypothesis proposed in Chapter 2. A secondary goal of the study was to identify feedstock-specific enzymes that may be of use in biofuel applications. For this purpose, we also assayed enzyme activity on model substrates using zymograms. The specific goals of this Chapter were to (1) evaluate the effect of diet on relative abundance of gut prokaryotic populations of *M. darwiniensis* that were fed on wood (pine and eucalyptus) and plant substrates (sugarcane mulch and cotton) at multiple time points during the feeding trial, and (2) to identify enzymes with high activity on specific feedstocks of interest.

3.2 Methods

3.2.1 Rearing of termites under laboratory conditions

Feeding experiments were performed over a period of 14 months (from December 2012 to March 2014) in the insectary at the School of Biological Sciences, the University of Queensland (UQ). *M. darwiniensis* (Figure 3.1) were collected by the Department of Primary Industry and Fisheries, Darwin (Thistleton laboratory) and live termites were shipped to UQ in 232 x 142 x 64 mm containers containing *Eucalyptus regnans* saw dust and vermiculite to minimise stress to termites by providing cushion during transport (Howick et al., 1975). Shipping took two to three days and upon arrival, termites were visually inspected to ensure that they were healthy. This was important as shipping can cause termite to experience stress and may cause them to develop disease.

In the laboratory, containers were emptied onto a large tray to separate termites from saw dust and vermiculite. Termite individuals were carefully picked up using a pair of fine tip tweezers. Termites were sorted into different castes (mainly three group; soldiers, workers and other castes) and only healthy, freely moving workers were selected for experiment.

3.2.1.1 Optimising maintenance conditions for feeding experiments

In the laboratory, termites were divided into two batches. The first batch of *Mastotermes* was reared in 165 x 165 x 191 mm containers with *E. regnans* blocks. Container lids were replaced with metal mesh wires (6.5 x 6.5 x 0.6 mm) to create ventilation. The second batch of *Mastotermes* was reared in 232 x 142 x 64 mm containers after the first batch of individuals died from desiccation. Termites were kept in a tall standing cabinet with ventilation and were maintained in the dark at 27±2°C. The first batch of *Mastotermes* died within three days. The second batch of *Mastotermes* was monitored for a period of two weeks with 100 µl of water added every two days. Humidity was measured using digital hygrometer.



Figure 3.1: Photos of a (top) worker *M. darwiniensis* specimen and (below) dissected *M. darwiniensis* worker whole gut.

3.2.1.2 Feeding experiments

For feeding experiment 1, upon arrival in the laboratory, three tubes of five workers each were subsampled from the colony for DNA extraction and 50 workers for proteome analysis, and were stored at -80°C till further processing. Similar to pre-trials, the remaining termites were separated from saw dust and vermiculite and divided into nine 232 x 142 x 64 mm containers and labelled accordingly (**Table 3.1**). Experiment was conducted in three replicates on three diets: *E. regnans*, hay and cotton. Each container was supplied with 100 μl of water and refilled every two days. All containers were kept at room temperature in a dark ventilated cabinet for period of four days as described above. *Mastotermes* workers reared on *E. regnans* and cotton were collected and stored at -80°C at day four till further processing. Individuals in the hay treatment were infested with fungus and discarded, consequently feeding experiment was halted.

For feeding trial 2, upon arrival in the laboratory, collection of *Mastotermes* at 0 hr were repeated as of feeding trial 1 (**Table 3.1**). The remaining termites were separated into 12 232 x 142 x 64 mm containers with ~200 workers and one to two soldiers per container. All termites were first placed on *E. regnans* for three days to allow them to adapt to the laboratory conditions. Different feedstocks (*E. regnans*, *Corymbia citrodora* subsp. *variegata*, sugarcane mulch, hay and cotton) were tested in three replicates. *C. citrodora*, genetically modified eucalyptus plant biomass was used as second generation biofuel feedstock of interest (Lupoi et al., 2015). Hay and sugarcane mulch were used as grass diets. Each container was supplied with 100 μl of water and refilled when required. All containers were maintained in conditions as described above. *Mastotermes* workers were collected from each container at day four and day seven into a 2 ml tube, and stored at -80°C till further processing.

3.2.2 Microbial community profiling of the gut of *M. darwiniensis*

Frozen specimens were thawed on ice, and gut tracts were extracted using clean sharp tweezers. The guts were immediately transferred into a sterile 1.5 ml tube on ice and subjected to DNA extraction. Total genomic DNA was extracted from five pooled whole gut samples using FastDNA® SPIN kit for Soil (MP Biomedicals, Australia). Termite guts were added to a lysing matrix, treated with lysis buffer, and underwent bead beating in the Vortex-Genie® 2 (MoBio Laboratories, USA). DNA was bound to silica matrix and washed and eluted in DNase-free water. DNA quality was evaluated using gel electrophoresis on 1.0% agarose gels stained with SYBR Safe, and visualised on a CCD compact image system (Major Science, USA). DNA quantification was carried out using the Qubit™ fluorometer and QuantIT ds-DNA BR assay kit (Invitrogen, Australia). DNA concentration varied depending on the biomass of the whole gut. DNA concentrations were standardised across all samples to 20 µg/ml, diluting (where necessary) with Ultrapure™ distilled water (Invitrogen, Australia).

3.2.3 PCR amplification and amplicon sequencing

For the feeding trial 1 samples, DNA were prepared, purified and sequenced according to methods described in Chapter 2. For the feeding experiment 2 samples, the universal primer pair 926F (5'-AAACTYAAAKGAATTGRCGG-3') and 1392R (5'-ACGGGCGGTGWGTRC-3') was used to amplify the V6 to V8 variable regions of the SSU rRNA gene (Lane, 1991). Primer sequences were modified by incorporation of Illumina specific adapter sequence (i5: TCGTCGGCAGCGTC; i7: GTCTCGTGGGCTCGG). PCR products of ~466 bp were amplified in standard PCR conditions using workflow outlined by Illumina (#15044223).

To ensure that similar numbers of sequencing reads were obtained for each sample, PCR amplicons were pooled in equal concentrations after amplification and then purified using the Agencourt® AMPure® XP Kit (Beckman, USA). Amplicons were quantified with the Qubit™ fluorometer and QuantIT ds-DNA BR assay kit. Purified amplicons were indexed with unique 8 bp barcode using the Illumina Nextera Kit A-D (Illumina FC-131-1002) to identify amplicons originating from different samples in the same sequencing reaction. Purified, pooled and barcoded amplicons were paired end sequenced with both the forward and reverse primers on Illumina MiSeq (Illumina, San Diego, CA, USA) with V3 300 bp chemistry according to manufacturer's protocol.

Table 3.1: Workers collected at multiple time points from feeding trial 1 and 2.

	Day 0	Day 1	Day 4	Day 7
Feeding Trial 1				
<i>Eucalyptus regnans</i> (Control)				
ER1	5	5	5	-
ER2	5	5	5	-
ER3	5	5	5	-
Cotton				
CT1		5	5	-
CT2		5	5	-
CT3		5	5	-
Feeding Trial 2				
<i>Eucalyptus regnans</i> (Control)				
ER1	5 [^]	-	5 [^]	5
ER2	5 [^]	-	5 [^]	5
ER3	5 [^]	-	5 [^]	5
<i>Corymbia citrodora</i>				
CC1		-	5 [^]	5
CC2		-	5 [^]	5
CC3		-	5 [^]	5
Sugarcane mulch				
SC1		-	5 [^]	5
SC2		-	5 [^]	5
SC3		-	5 [^]	5
Cotton				
CT1		-	5 [^]	5
CT2		-	5 [^]	5
CT3		-	5 [^]	5

[^]Additional 50 workers collected for proteomics

3.2.4 Sequence processing and analysis

SSU rRNA sequence data obtained from the Illumina MiSeq run were demultiplexed and analysed using Quantitative Insights Into Microbial Ecology (QIIME) (Caporaso et al., 2010). Reads were trimmed to 200 bp and clustered into operational taxonomic units (OTUs) with a threshold of 97% sequence similarity using MCL (Kunin and Hugenholtz, 2010). OTU representatives were compared to the Greengenes database (May 2013 release) for taxonomy assignment using BLASTn (Camacho et al., 2009). Termite host OTUs were removed and the number of sequences were normalised to 14,000 sequences per samples for sampling depth. Normalised OTU table were adjusted to account for differences in SSU rRNA copy numbers between taxa using CopyRighter (Angly et al., 2014). Observed species (richness) and Simpson index were calculated for each

sample using QIIME. A table which lists the relative abundance of each OTU in each sample was generated. Differences in the composition of *Mastotermes* gut communities at different diet and time points were assessed using permutational multivariate analysis of variance (PERMANOVA) and principal component analysis (PCA). All analyses were conducted and visualised using R version 3.2.0 with the vegan package (Oksanen et al., 2007).

3.2.5 Proteomics

3.2.5.1 Preparation of crude enzyme extract

Frozen termite specimens were thawed on ice, and gut tracts were extracted using clean sharp tweezers. The guts of 20 workers were immediately transferred into a sterile 1.5 ml tube with 150 µl of phosphate buffered saline (PBS, pH 7.4) on ice. Proteinase inhibitor solution (1X Complete Mini EDTA-free; Roche) was added to the extraction buffer to prevent protein degradation. The gut samples were homogenised with a sterile pellet pestles, vortex-mixed and centrifuged at 14,000 g for 10 min at 4°C. The supernatants were pooled into a new sterile 1.5 ml tube and stored at -80°C. The obtained supernatant is referred as crude enzyme extract. Three replicates were prepared for each feedstock in the same manner.

3.2.5.2 Measurement for cellulase activity

Total protein concentration in crude enzyme extract was measured using Bradford method (Bradford, 1976) via Coomassie Protein Assay Kit using bovine serum albumin (BSA) as a protein standard (working concentration of 100-1500 µg/ml). Absorbance was measured at 595 nm with a plate reader. Activity of endoglucanase (CMCase) were measured from *M. darwiniensis* gut content enzyme extracts based on the release of reducing sugar using the dinitrosalicylic acid (DNS) assay (Miller, 1959; Coughlan and Moloney, 1988). All crude enzyme extract were diluted to a working concentration of ~1.2 µg/ml in 50 mM citrate-phosphate buffer (pH 6.5). The 96-well microplate protocol was performed as previously described (Xiao et al., 2005; dos Santos Castro et al., 2014). Briefly, 30 µl of crude enzyme extract was combined with 30 µl of substrate (1% carboxymethyl cellulose (CMC) in 50 mM citrate-phosphate buffer, pH 6.5). The reaction was incubated at 50°C for 30 min in PCR thermocycler and then stopped by addition of 60 µl of DNS reagent. The colour reaction was developed at 95°C for 5 min, followed by cooling at 4°C for 1 min and hold on 20°C. A 100 µl aliquot of each reaction was transferred to the flat-bottom 96-well plate and absorbance was measured at 540 nm using SpecMAx 190 microplate reader. Glucose standards of 0, 0.3, 0.6, 0.9, 1.2 and 1.5 mg/ml were prepared in 50 mM citrate-phosphate buffer pH 6.5 and used for the standard curve construction. Enzyme controls were prepared by adding 30 µl of crude enzyme

extract to the pre-incubated (50°C for 30 min) substrate and DNS was added to halt enzyme activity. Each assay was performed in triplicate along with glucose standard and controls. One unit of the cellulolytic activity was defined as the amount of enzyme capable of releasing 1 μ M of glucose per minute.

3.2.5.3 Protein electrophoresis and zymogram analysis

Cellulase activities of crude extract from *M. darwiniensis* gut content were detected by performing sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) through in-gel substrate and/or overlay zymogram techniques. A commercial *Aspergillus niger* cellulase (Sigma–Aldrich, Australia) was used as a positive control in all tests. For in-gel substrate zymograms, samples were mixed with Laemmli sample buffer (Laemmli, 1970), heated to 95°C for 3 min and loaded onto 4-12% SDS-PAGE gel containing 0.1% CMC. Sample loading pattern (20 μ g/lane) was repeated onto the same gel to produce identical gels. Electrophoresis was carried out at constant voltage (100 V) for 150 min at room temperature. After electrophoresis, the gel was then cut into half, where the first half was used for zymograms and second half for protein visualisation by staining with Coomassie brilliant blue R-250. For zymograms, the gel was washed four times for 15 min each in 25% (v/v) isopropanol followed by four washes 15 min each in 50 mM citrate-phosphate buffer pH 6.5 at room temperature. The gel was then incubated at 50°C for 30 to 60 min in 50 mM citrate-phosphate buffer pH 6.5, washed with water and stained with 0.1% Congo red for 30 min. Excess Congo red was removed and the gel was washed with 1 M NaCl and 10 min with 0.1 M Tris-HCl pH 8.0 until clear bands became visible indicating cellulase activity. To increase visualisation of the activity bands, 5% acetic acid was introduced into the incubation chamber with gel. The gel was visualised and imaged under ultraviolet light and aligned with the Coomassie blue stained gel. The zymogram gel was documented using the Odyssey Infrared Imaging System (Li-Cor Bioscience).

Overlay zymograms was prepared on a 8-16% Tris-Glycine Mini Protein gel (Novex, Invitrogen) using Novex SDS sample buffer (2X) without a reducing agent and sample heating. Samples (20 μ g/lane) were loaded in the same manner as for in-gel substrate zymograms. The gel was run at constant voltage (120 V) for 190 min at 4°C using pre-chilled running buffer. Following electrophoresis, the gel was cut into half as described above. The zymogram half of the gel was washed four times for 15 min each in cold 25% (v/v) isopropanol followed by four washes 15 min each in cold 50 mM citrate-phosphate buffer pH 6.5. The polyacrylamide gel was then placed on the top of a 2% agarose gel made in 50 mM citrate-phosphate buffer pH 6.5 containing 0.1% CMC. The assembled gels were incubated at 50°C for 30 to 60 min in humidity chamber (container with wetted paper towel). The agarose gel was stained and visualised as described above.

3.2.5.4 Protein sequencing and identification

The bands of interest correlated to activity zones in zymograms were carefully excised from the reference gel (Coomassie blue stained gel) and transferred to 1.5 ml centrifuge tubes for trypsin digestion. Briefly, gel pieces were washed twice for 30 min each with 25 mM ammonium bicarbonate in 50% acetonitrile, reduced with 10 mM DDT in 25 mM ammonium bicarbonate for 45 min at 55°C and then alkylated with 50 mM iodoacetamide in 25 mM ammonium bicarbonate for 30 min at room temperature in the dark. To remove iodoacetamide, gel pieces were washed three times for 10 min each with 25 mM ammonium bicarbonate, followed by dehydration with 100% acetonitrile. Trypsin digestion of dehydrated and air dried gel pieces was performed at 37°C overnight using sequencing grade modified trypsin (Promega, Madison) in 25 mM ammonium bicarbonate. The peptides were extracted with 1% trifluoroacetic acid in 50% acetonitrile, and dried using a Speed Vac. Peptide samples were cleaned up using ZipTip pipette tips (Millipore) according to the manufacturer's protocol with some modifications prior to mass spectrometry analysis. Briefly, ZipTip tips were equilibrated with 100% acetonitrile, followed by peptide binding with 0.1% trifluoroacetic acid in 5% acetonitrile, and finally peptides were eluted into 0.1% trifluoroacetic acid in 80% acetonitrile. Eluted samples were further diluted 10 times with sterile water and a 100 µl was submitted for peptide sequencing in the HPLC grade Agilent vial. Peptide sequencing and analysis were performed at the Mass Spectrometry Research Facility at the School of Chemistry and Molecular Bioscience, UQ, on a TripleTof 5600 instrument (ABSciex) (liquid chromatography–mass spectrometry/mass spectrometry (LC-MS/MS)).

All MS/MS spectra were analysed using Mascot (Matrix Science, London UK) (provided by the Australian Proteome Computational Facility, <http://www.apcf.edu.au/>) to search against Swiss-Prot and/or LudwigNR databases, using the following parameters; (1) species: all species, (2) fixed modifications: carbamidomethyl of cysteine, (2) variable modifications: oxidation of methionine, (4) enzyme: trypsin, (5) number of allowed missed cleavages: up to 2, (6) peptide (MS) mass tolerances: 50 ppm and (7) MS/MS mass tolerance: 0.1 Da. Peptides were also identified against our *M. darwiniensis* metagenome database (described in detail in Chapter 4) by translating the nucleotide to amino acid sequences. Search was performed using Mascot search algorithm with the above-mentioned parameters. Identified peptide sequences based on the metagenome data were assigned using protein Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) algorithm, against non-redundant protein sequences (nr) database.

3.3 Results and Discussion

3.3.1 Optimisation of rearing termites under laboratory conditions

Termites are known to be sensitive to variable humidity which can impact on various aspects of colony life including individual and colony survival (Fuller et al., 2011; Fuller and Postava-Davignon, 2014). As termites are vulnerable to desiccation, they construct their nests or mounds in a way that maintains a constant humidity (60-75%) independent of external conditions (Davenport, 2012). Therefore, we tested container size and ventilation over a period of two weeks before beginning our feeding experiments.

Evaluation of containers for termite survival. Termites were initially trialled in 165 x 165 x 191 mm plastic containers with metal mesh replacing the lid for ventilation. Five to six hundred *Mastotermes* workers placed in these containers all died over a period of four to seven days from desiccation due to the large ventilation opening resulting in too low humidity (50%). Mesh openings were therefore reduced in size to a 10 mm diameter opening in the lid of a smaller container measuring 232 x 142 x 64 mm. Two to three hundred workers were successfully maintained at 60-75% humidity in these containers but only up to two weeks before population health declined. This is consistent with declining health and survival rate of termite individuals in laboratory kept small colonies as previously reported for *Nasutitermes* (Husseneder et al., 2010) and *Coptotermes* (Tanaka et al., 2006).

Feeding trial 1. *Mastotermes* were reared on three feedstocks, *Eucalyptus regnans*, hay and cotton, in the 232 x 142 x 64 mm with 10 mm ventilation grids, using three replicated containers per feedstock for a planned period of seven days (**Figure 3.2** and **Table 3.1**). Previous feeding studies involving termites suggest that seven days is sufficient time for the gut community to adapt to a new feedstock (Husseneder et al., 2009; Boucias et al., 2013; Raychoudhury et al., 2013; Sethi and Scharf, 2013). *E. regnans* was used as a control as this was the biomass that the termites were reared on by our collaborators in the Northern Territory. Hay was used as an example of a C3 grass diet which has different hemicellulose composition to wood (Knobbe et al., 2006), and cotton was used as a pure crystalline cellulose control. To allow the termites to adapt to the insectary conditions, they were left undisturbed in their original container for two days before separating them into sets of ~200-300 individuals per container. Visual observation after one, two, and four days revealed that termites on *E. regnans* and cotton were actively feeding on the supplied substrates, however, their counterparts on hay were infested with fungus by day four and the experiment was suspended at this point. The rapid mortality observed on hay was due to fungal

spoilage of the hay at the experimental humidity of 60-75% (Knapp et al., 1975). Therefore, hay was replaced by sugarcane mulch as an alternative (C4) grass in feeding trial 2.

Feeding trial 2. *Mastotermes* workers were reared on *E. regnans*, *Corymbia citrodora*, cotton and sugarcane mulch using three replicated containers per feedstock for a period of seven days (**Table 3.1**). *C. citrodora* was included as a biofuel feedstock of interest. Unexpectedly, individuals in all the containers became infested with mites after two days (**Figure 3.3**), which quickly led to 100% mortality. Mites are known parasites of termites (Weiser and Hrdy, 1962). The presence of mites attached to various body parts, particularly the head (**Figure 3.3**), indicate that the termite colony is undergoing stress (Korb and Fuchs, 2006), suggesting that social grooming did not happen or was less frequent (Micheal Neal, personal communication). Social grooming is an important defensive behaviour of termites against parasites such as mites and other parasites (Korb and Fuchs, 2006). Since we were unable to identify a cause for the mite infestation, we simply repeated the feeding trial with a fresh batch of *Mastotermes* workers. Individuals in all containers appeared healthy and were actively feeding and producing frass over the seven day feeding trial.

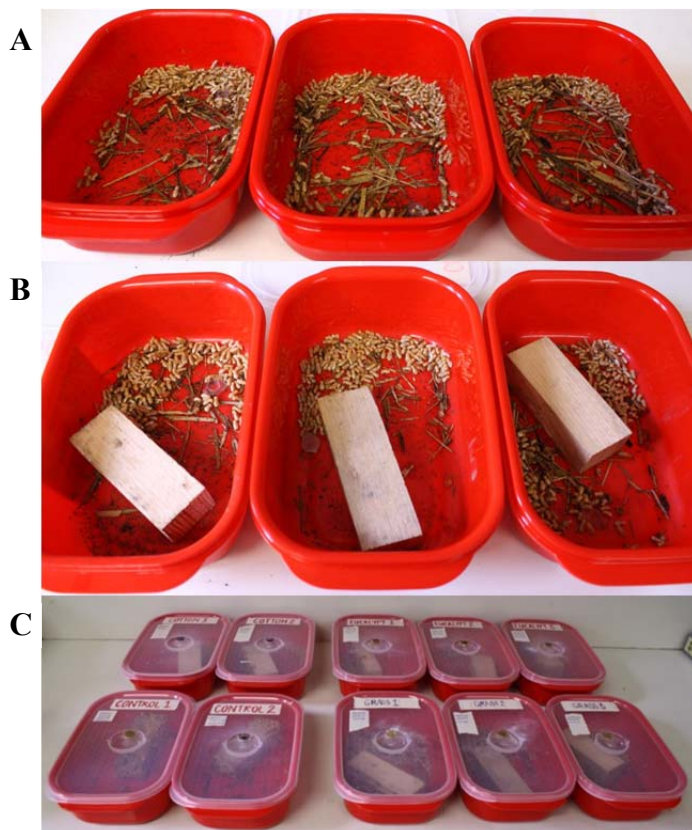


Figure 3.2: Photos of 232 x 142 x 64mm containers (A) with sugarcane mulch, (B) *Eucalyptus* wood block, (C) set up of feeding experiment.

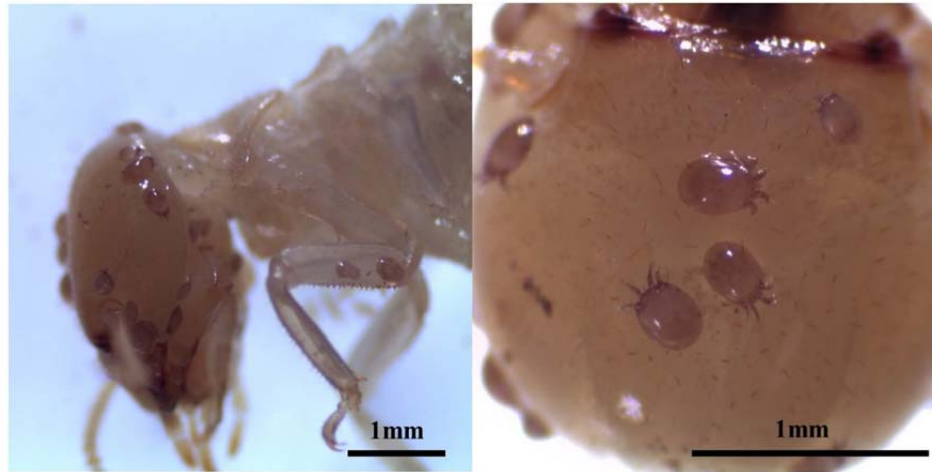


Figure 3.3: Photos of worker *M. darwiniensis* (left) infested with mites and (right) a close up of mite on the head. Scale bar = 1 mm in both panels.

3.3.2 Effect of different lignocellulose feedstocks on gut microbiota of *Mastotermes*

3.3.2.1 *Mastotermes* gut community profiling

Whole guts were removed and pooled from five workers collected at multiple time points during the two feeding trials (**Table 3.1**). Culture-independent microbial community profiles were determined via SSU rRNA gene amplicon sequencing using the primer pairs 926F and 1392R that broadly target all three domains of life (Lane, 1991). The Roche 454 Pyrosequencer at the Australian Centre for Ecogenomics was decommissioned between the first and second feeding trials, consequently, pyrosequence reads (pyrotags) were obtained for the first study, and MiSeq Illumina reads (mitags) were obtained for the second study. A total of 47,557 quality trimmed pyrotags were produced from the 15 samples of the first feeding trial ranging from 724 to 5,258 per sample after removal of termite host SSU rRNA gene sequences, which comprised 16 to 56% of total reads for each sample. Specimens were randomly resampled to a depth of 700 reads. A total of 1,525,235 quality trimmed mitags were produced from the 27 samples of the second feeding trial ranging from 17,935 to 134,301 per sample after removal of termite host SSU rRNA gene sequences, which comprised from 23 to 33% of total reads for each sample. Specimens were randomly resampled to a depth of 14,400 reads. To assess the completeness of the diversity of microbial taxa in termite guts that were sampled in the feeding experiments, we performed rarefaction analyses. Rarefaction analysis suggested that the gut microbiota of *Mastotermes* in feeding trial 2 were more diverse than those in feeding trial 1 (Appendix B: **Figure S3.1**). Although the alpha diversity can be significantly affected by sequencing length and depth, these biases can be corrected by adjusting to a suitable cutoff for low abundance OTU filtering (Tremblay et al., 2015). The greater alpha diversity observed in mitag reads as compared to pyrotag indicates that this sequencing platform would be

favoured for rare biosphere applications (Tremblay et al., 2015). In addition, rarefaction analysis of the average replicates in the gut samples indicated a larger richness in day four gut samples. The 15% dropped in the number of observed OTUs in the day seven gut samples may indicate a decrease in diversity with time. The resampled data was normalised for SSU rRNA copy number variation using CopyRighter which can vary by up to an order of magnitude between prokaryotic genera (Angly et al., 2014). The majority of the overall non-host amplicons reads from the whole *Mastotermes* gut samples were bacterial (87.7-89.6% on average) followed by archaea (~10.0% on average) and protists (0.5-2.2% on average) (Appendix B: **Table S3.1**).

3.3.2.2 Variation of gut microbial profiles of *Mastotermes* using different sequencing platforms

Since we were obliged to change sequencing platforms, a primary question was whether the pyrotag and mitag data were comparable. In order to address this, we first assessed the biological differences between samples within the same treatment. Many of the biological replicates showed appreciable differences but did not differ significantly ($P = 0.833$) compared to different treatments (Appendix B: **Table S3.1**). As previously noted in the community profiling survey presented in Chapter 2, the variation observed between biological replicates suggests greater inter-individual gut microbiome differences than anticipated assuming that technical replication is sound (Appendix A: **Figure S2.3**). Inter-individual gut microbiome differences from the same colony have previously been reported (Minkley et al., 2006; Boucias et al., 2013). Despite this variation and the use of two different sequencing platforms, gut community composition was comparable between the two feeding trials on *E. regnans* at time zero (**Figure 3.4**). Notable exceptions were the Bacteroidetes, Elusimicrobia and Spirochaetes (**Figure 3.4**). The Elusimicrobia had higher relative abundance in the mitag profiles while the reverse was true for the Bacteroidetes and Spirochaetes. To quantify the variance between the two datasets, we performed a permutational multivariate analysis of variance (PERMANOVA). The frequencies of 171 OTUs displayed variation ($P < 0.05$) between the two datasets suggest that sequencing platforms can account for up to 65% of the variance. Two major factors that can account for the observed differences are (1) methodological biases, such as error estimates and sequencing biases between the two sequencing platforms (Luo et al., 2012; Ratan et al., 2013), and (2) biological differences as observed between *Mastotermes* colonies collected from Darwin and Townsville (*see* Chapter 2 and 4). Since most published termite gut symbiont community profiling studies are based on 454 pyrosequencing (*see* **Table 1.3**), it is difficult to identify which sequencing platform provides greater accuracy of the *Mastotermes* gut profile based on comparison to previous reports alone. To this end, we compared gut profiles of two recent termite surveys by the same group that were conducted via 454 amplicon pyrosequencing (Dietrich

et al., 2014) and Illumina Miseq (Dietrich et al., 2014) pyrosequencing to further validate population differences between the two sequencing platforms. The three higher termite (*Nasutitermes corniger*, *Trinervitermes* sp. and *Cubitermes ugandensis*) gut profiles that were selected from both studies for the purpose of this analysis originated from the same colonies respectively. Similar to our results, the pyrotag- and mitag-based profiles from Dietrich et al. (2014) and Mikaelyan et al. (2015) were comparable with notable discrepancies in some populations (e.g. Acidobacteria, Bacteroidetes, candidate phylum TG3, Proteobacteria and Spirochaetes) (Appendix B: **Table S3.2**). We were unable to identify specific bacterial populations that were affected by the different sequencing platforms. This is due firstly to the use of different termite genera; in the Dietrich et al. (2014) and Mikaelyan et al. (2015) studies, higher termites were used and secondly the limitation of multiplex amplicon sequencing may lead to artificial inflation of microbial diversity (Kunin et al., 2010). Although the limitations for both the Illumina and 454 platforms have been extensively evaluated in terms of the sequencing error and artefacts in previous studies (Margulies et al., 2005; Erlich et al., 2008; Gomez-Alvarez et al., 2009; Quince et al., 2009), Roche 454 tends to produce a higher error rate, with up to 15% of generated sequences resulting from artificial amplification (Gomez-Alvarez et al., 2009; Luo et al., 2012). More importantly, it is well established that amplicon-based sequencing can be biased due to unequal amplification of SSU rRNA genes (e.g. overestimate of species diversity as a result of chimeric sequence formation) (Haas et al., 2011; Tremblay et al., 2015). There is therefore a need for a less biased amplification approach such as shotgun sequencing-based community profiling to reduce limitations that exist with current amplicon-based sequencing methods (Darling et al., 2014; Hasan et al., 2014) (discussed in more detail in Chapter 4). Due to the uncertainty in comparing data from feeding trials 1 and 2, we focussed our attention on the second feeding trial which incorporated more feedstocks.

Taxon	pyrotag				mitag			
	C1_0_p	C2_0_p	C3_0_p	avg	C1_0	C2_0	C3_0	avg
Eukaryota;_Excavata	1.0%	0.5%	0.8%	0.7%	2.5%	3.3%	3.8%	3.2%
k_Archaea;p_Euryarchaeota	7.4%	6.8%	16.1%	10.1%	9.4%	9.1%	11.4%	10.0%
k_Bacteria;p_Acidobacteria	0.2%	0.0%	0.2%	0.1%	0.5%	0.5%	0.5%	0.5%
k_Bacteria;p_Actinobacteria	1.7%	1.3%	2.3%	1.8%	1.9%	1.9%	2.0%	2.0%
k_Bacteria;p_Bacteroidetes	52.8%	55.4%	42.5%	50.2%	39.4%	36.6%	33.5%	36.5%
k_Bacteria;p_Elusimicrobia	8.3%	6.8%	4.6%	6.5%	18.3%	15.3%	15.2%	16.3%
k_Bacteria;p_Firmicutes	14.5%	16.2%	22.5%	17.7%	14.3%	22.0%	20.6%	19.0%
k_Bacteria;p_Fusobacteria	4.8%	4.4%	4.3%	4.5%	4.8%	4.3%	4.4%	4.5%
k_Bacteria;p_Planctomycetes	0.3%	1.0%	0.4%	0.6%	0.6%	0.9%	0.9%	0.8%
k_Bacteria;p_Proteobacteria	1.8%	2.0%	1.1%	1.6%	4.2%	2.3%	4.2%	3.6%
k_Bacteria;p_Spirochaetes	6.2%	5.0%	4.3%	5.2%	2.2%	2.4%	2.2%	2.3%
k_Bacteria;p_Synergistetes	0.7%	0.6%	0.5%	0.6%	0.3%	0.1%	0.3%	0.2%
k_Bacteria;p_TM7	0.1%	0.0%	0.1%	0.1%	0.5%	0.3%	0.3%	0.3%
k_Bacteria;p_Tenericutes	0.1%	0.1%	0.1%	0.1%	0.3%	0.4%	0.5%	0.4%
k_Bacteria;p_Verrucomicrobia	0.2%	0.0%	0.2%	0.1%	0.7%	0.5%	0.2%	0.5%

Figure 3.4: Relative abundances ($\geq 0.2\%$) of *M. darwiniensis* gut profiles on the *E. regnans* control diet using different sequencing platforms. Each row represents a microbial taxon with relative abundance indicated by red shading and taxa with notable differences between the two platforms indicated by yellow.

3.3.2.3 Effect of time and diet on microbial gut community profiles

Phylum-level groups identified in the gut microbiome at three time points (0 hr, Day 4 and Day 7) were largely consistent with notable differences in relative abundance of some phyla between termites at time points four and seven (**Figure 3.5**). The relative abundance of core termite gut bacterial taxa; Bacteroidetes, Spirochaetes and Proteobacteria, present in all profiled samples in this study, were fairly consistent across all samples which further supports their essential roles in the gut community despite differences in diet. Four prokaryotic phyla, Euryarchaeota, Elusimicrobia, Actinobacteria and Firmicutes showed pronounced differences between samples at day four and day seven, irrespective of diet. The Euryarchaeota and Elusimicrobia are more abundant at day four in comparison to day seven. The opposite is true for Actinobacteria and Firmicutes. Most members of Elusimicrobia and Euryarchaeota are intracellular bacterial symbionts of flagellates in the gut of lower termites and wood-feeding cockroaches (*Cryptocercus punctulatus*) (Ohkuma, 2008; Strassert et al., 2012), although a small number of Euryarchaeota such as some *Methanobrevibacter* are attached to the gut epithelial wall. Although only a relatively small percentage of *Mastotermes* protists were detected in this study as an artefact of the universal primers that do not target many microbial eukaryotes (see Chapter 2), it has been well demonstrated that flagellates make up the bulk of the microbial community in lower termites (Brune and Stingl, 2006; Tai et al., 2015). A declining trend in relative abundance of detected protists was observed between day four to day seven in all feeding treatments (**Figure 3.5**) consistent with the observed reduction in Elusimicrobia and Euryarchaeota. An opposite trend was observed for the Actinobacteria and Firmicutes where

the relative abundance increases from day four to day seven in all diets. Similar shift patterns of these dominant bacterial phyla, specifically Elusimicrobia and Actinobacteria, were previously observed in a closely related wood-feeding roach, *C. punctulatus* (Berlanga et al., 2009). Berlanga et al. (2009) reported that the relative abundance of Elusimicrobia decreases and Actinobacteria and Firmicutes increase significantly after fasting. This suggests that at least some of the observed community composition changes may have been due to stress related to the change in diet, rather than strict adaptation to a new feeding substrate.

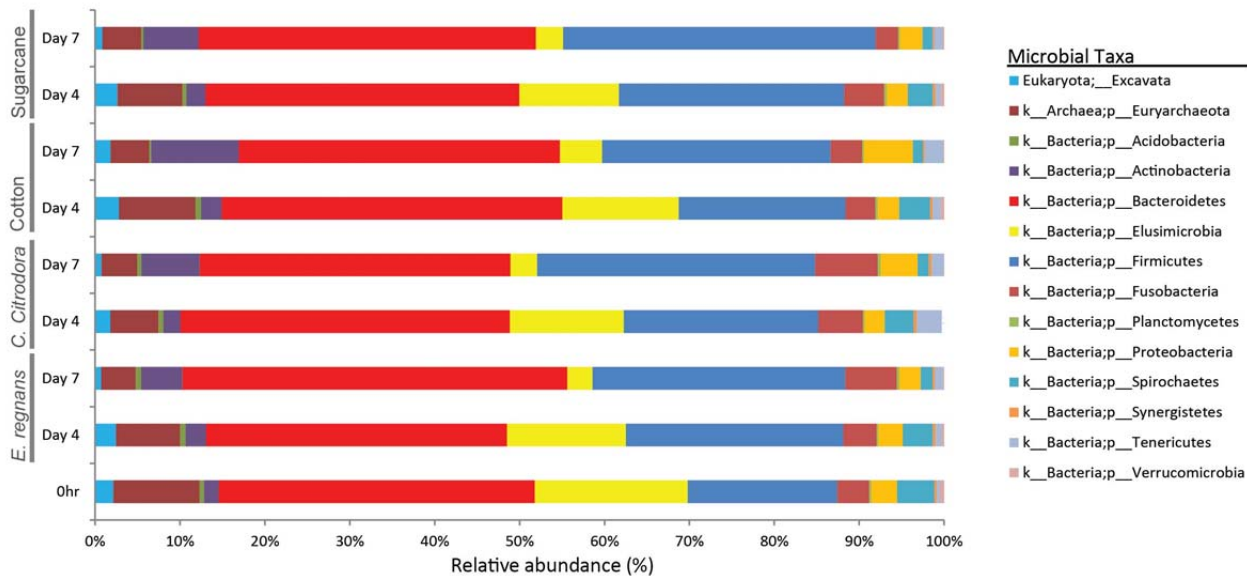


Figure 3.5: Phylum-level summary of microbial taxa between wood- and grass-fed *Mastotermes*. The bar graph represents the averaged of three replicates per treatment.

Microbial gut community composition was analysed in greater detail using 97% identity operational taxonomic units (OTU) (**Figure 3.6** and Appendix C: **Table S3.1**). Two *Methanobrevibacter* OTUs (OTU9 and OTU38) were more abundant at day four than day seven for all diets. *Methanobrevibacter* sp. are found either in the periphery of the hindgut or as an endosymbionts of protists which suggests that these OTUs are associated with declining gut protist species. It has been suggested that this association is a strategy in maintaining an anoxic condition in the termite gut ecosystem (Brune, 2011). Other likely protist symbionts and protist OTUs also declined in relative abundance between day four and seven. These included the dominant *Endomicrobia* OTU (OTU27) on day four (**Figure 3.6**) which has 100% identity to an uncultured *Endomicrobia* endosymbiont of cellulolytic *Deltotrichonympha* sp. in the Parabasalia group (Li et al., 2003; Watanabe et al., 2006; Ikeda-Ohtsubo et al., 2007) and three parabasalid OTUs (OTU131, OTU77 and OUT161). These parabasalid OTUs are affiliated to *Deltotrichonympha* sp. (OTU131; 0.2 – 3.8%), *Koruga* sp. (OTU77; 0.1-0.4%) and *Metadevescovina* sp. (OTU161; 0.1-0.5%). In a feeding trial of the lower termite *Coptotermes formosanus*, Tanaka et al. (2006) demonstrated that changes in diet resulted in

changes in the protist community. Because the study compared the gut profiles at the end of 30 days, the differences in the number of protists were only compared between individuals on different diets and possible changes that may have been due to stress over time were not recorded. Findings from this study showed a strong correlation between protists on wood and artificial diets. Individuals on wood (complex lignocellulose) substrates had high abundance of protists as compared to those on artificial diets (cellulose, cellobiose and glucose). This reduction in relative abundance of protists provides evidence that they are the key players in lignocellulose digestion, particularly of cellulose and hemicellulose. These authors concluded that the protist community is unable to effectively use the artificial substrates resulting in a decrease in most protist populations (Tanaka et al., 2006). This may also be the case in feeding trial 2 for the cotton and sugarcane mulch substrates.

For the inclining trend observed in some of the prokaryotic populations, five *Actinomycetales* OTUs (OTU51, OTU55 and OTU147-*Microbacterium* sp., OTU52-*Corynebacterium* sp., and OTU42-*Tsukamurella* sp.) mainly accounted for the higher abundance of Actinobacteria present in gut profiles at day seven (**Figure 3.6** and Appendix B: **Table S3.1**). Interestingly, most of these OTUs are associated with aerobic bacteria suggesting that the gut condition was favouring the growth of these populations. Two Firmicutes OTUs, OTU69 and OTU70, identified as *Leuconostocaceae* sp. and *Lachnospiraceae* sp., respectively, increased in number from day four to day seven in all treatments. *Leuconostocaceae* sp. were previously reported in lower termite genera *Mastotermes*, *Hodotermes*, *Reticulitermes* and *Coptotermes* in low abundance (Dietrich et al., 2014; Butera et al., 2015) but were relatively high in the *Mastotermes* profiles at day four and day seven. Members of the family *Leuconostocaceae* are generally involved in carbohydrate fermentation isolated from food goods ranging from vegetables, fruits, fish and dairy products to spoiled refrigerated meats (Robinson and Batt, 1999), yet their roles in the termite gut are unclear. *Lachnospiraceae* sp. have previously been reported in the gut periphery of lower and higher termites (Thompson et al., 2012) and were consistent with the appearance of aerobic Actinobacteria at day seven. Dissolution of physiochemical gradients, including oxygen, have been noted in cockroach guts when the protist community declines (Berlanga et al., 2009). This could explain the observed population shifts between days four and seven, i.e. decrease in protist and protist symbiont OTUs and increases in putatively aerobic bacterial populations. These shifts may point to a stress response to changes in diet, but since such shifts were also observed in the *E. regnans* controls, smaller colony size and container conditions may have contributed to the stress response.

Two other important OTUs that were abundant in all *Mastotermes* samples despite the compositional shifts observed in some phyla were OTU57 (~10%) and OTU97 (~4%) assigned to

Blattabacterium and *Fusobacterium* respectively (**Figure 3.6** and Appendix B: **Table S3.1**). *Mastotermes* are the only termite genus to retain the obligate fat body endosymbiont *Blattabacterium* from their ancestral cockroaches. Compared to *Blattabacterium* in omnivorous roach *Periplaneta*, those in *Mastotermes* and its sister wood roach *Cryptocercus* are smaller in genome size and has loss a fraction of genes involved in nutritional biosynthesis as a result of acquisition of gut microbes that are capable of supplying essential nutrients (Sabree and Moran, 2014). Even though *Blattabacterium* in *Mastotermes* are functionally diminished, its presence in the basal termite genus suggests that *Mastotermes* are lacking certain gut microbiome constituents required to fully take up the function of *Blattabacterium* (Sabree and Moran, 2014). Another gut symbiont, *Fusobacterium*, has been reported in almost all investigated *Mastotermes* gut profiles (3-6%) (Sabree et al., 2012; Dietrich et al., 2014; Sabree and Moran, 2014) (Appendix A: **Figure S2.8**). The only report on a potential role of *Fusobacterium* sp. is the member *Sebaldella termitidis*, belonging to the order *Fusobacteriales*, in lower termite *Reticulitermes* sp. were previously reported to produce acetic and lactic acid via sugar fermentation (Harmon-Smith et al., 2010) and *Fusobacterium* (OTU97) may perform a similar function in *Mastotermes*. Other OTUs that were consistent (>0.1%) across the different dietary treatment and time points were OTU113 (Acidobacteria), OTU66 (*Pirellulaceae*) and OTU13 (*Rhodocyclaceae*).

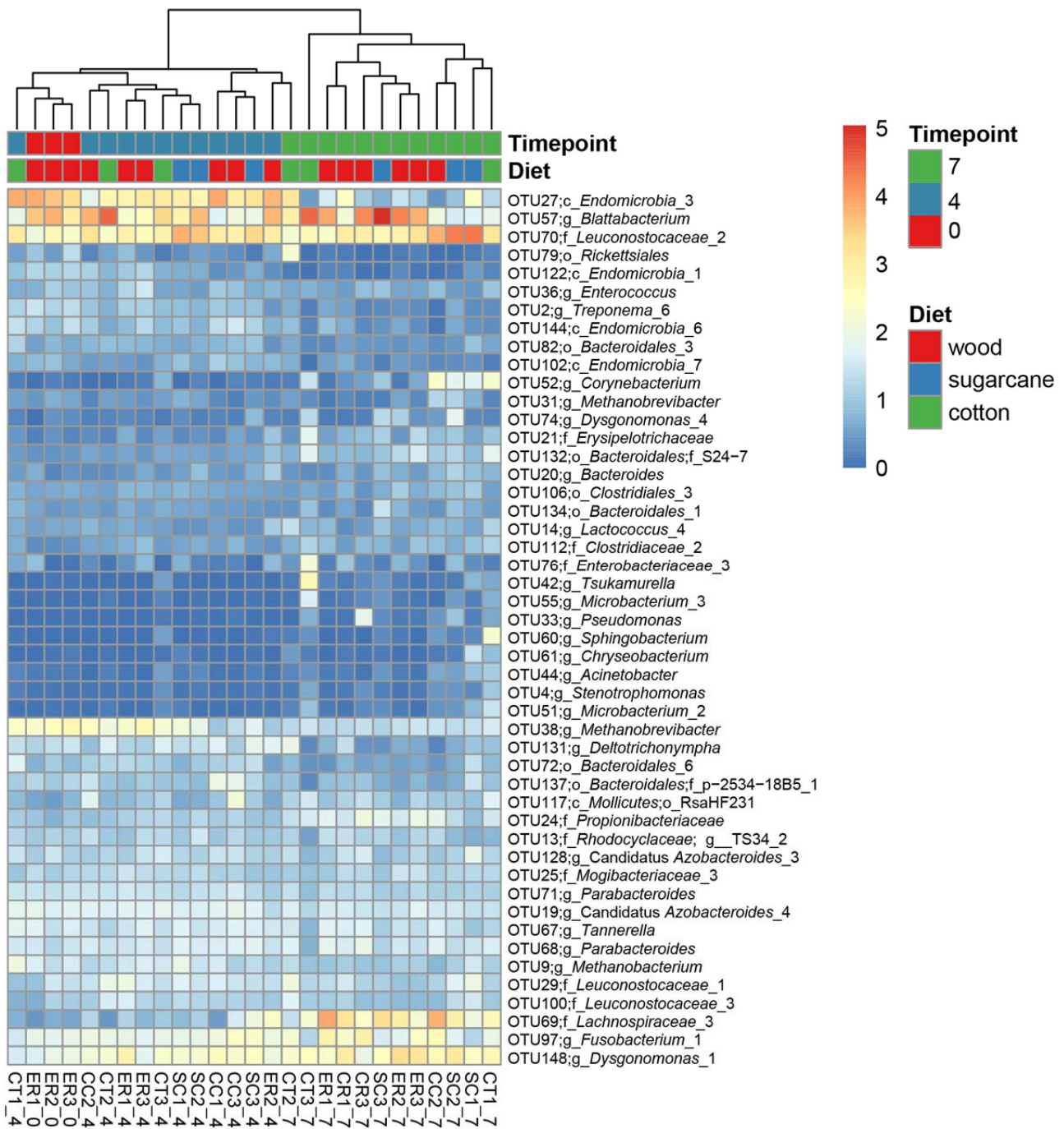


Figure 3.6: Heatmap showing microbial OTUs with $\geq 0.1\%$ relative abundance in the *Mastotermes* profiled gut samples in feeding trial 2. Each row represents an OTU with relative abundance indicated by shading according to the legend and each column represents a gut sample. The time point and diet for each sample is indicated at the top of the figure by colour according to the legend. Abbreviations; ER: *E. regnans*, CC: *C. citrodora*, CT: Cotton, SC: Sugarcane mulch.

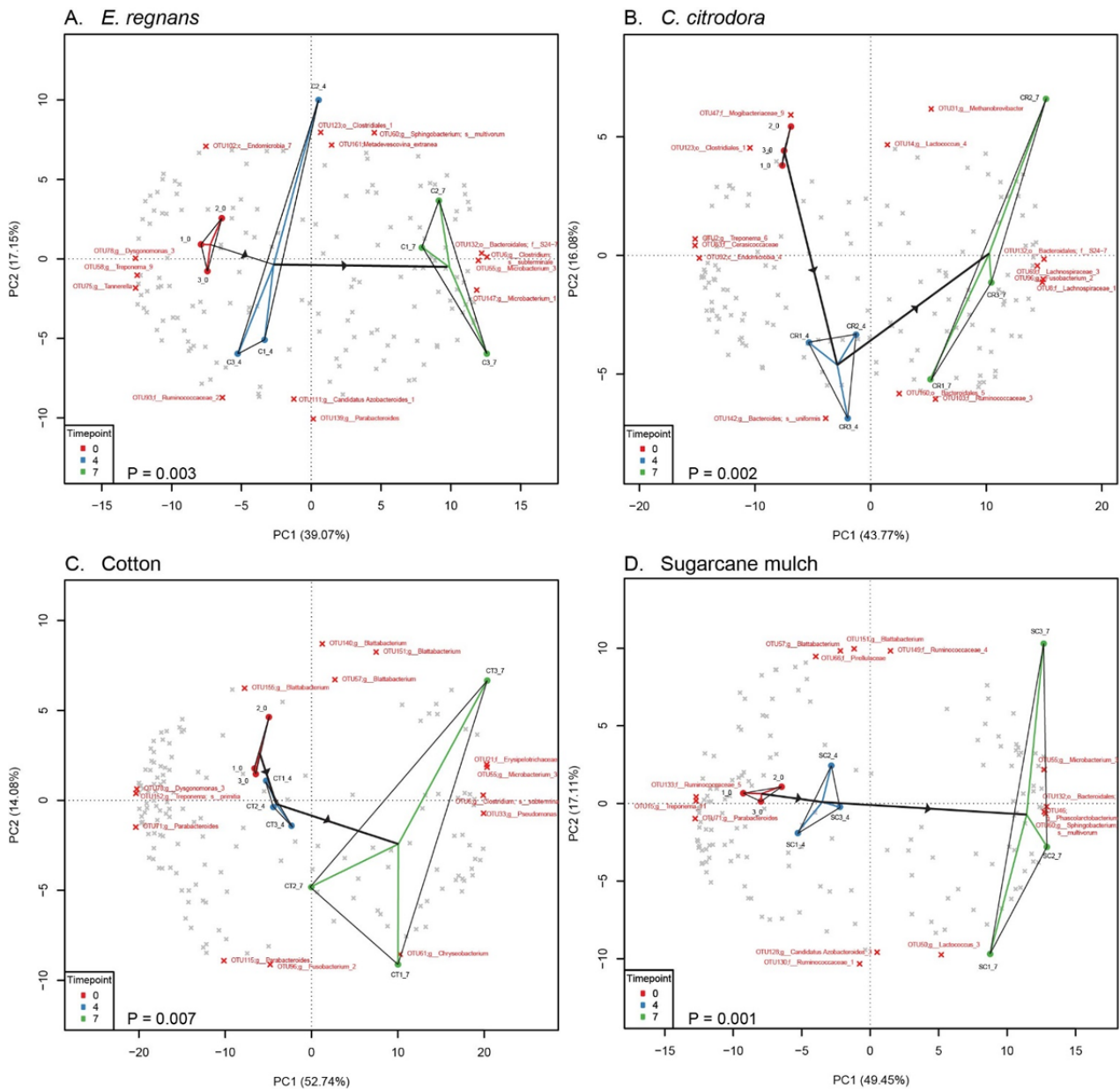


Figure 3.7: Principle component analysis (PCA) plots of microbial profiles obtained from *Mastotermes* gut samples biological replicates at 0 hr, Day 4 and Day 7. The colour indicates time points. Panels represent gut samples reared on different feedstocks; (A) *E. regnans*, (B) *C. citrodora*, (C) cotton and (D) sugarcane mulch. Principal components 1 and 2 explained (A) 39.07% and 17.15%, (B) 43.77 and 16.08%, (C) 52.74% and 14.08%, and, (D) 49.45% and 17.11% of the variance, respectively.

Since it is well established that there are associations between the protist-symbiont bacterial phyla to hindgut protists (i.e. decrease in Elusimicrobia and Euryarchaeota), protist-symbiont prokaryotes could provide a gauge for evaluating the presence and absence of protists in lower termites as observed in the community composition shifts from time zero to day seven (Ohkuma, 2008). The ability of microbes to adapt to changes in diet over the course of evolutionary timescales has been an open-ended research question. While termite feeding experiments have attempted to demonstrate that microbes in host-associated ecosystems such as observed in *Reticulitermes* and *Coptotermes* sp. are capable of adapting to dietary changes over a short period of seven days (Geib et al., 2008; Husseneder et al., 2009; Sethi et al., 2012; Raychoudhury et al., 2013), it is still unclear when the gut microbial community composition alters to accommodate a new diet. In a feeding study by Boucias et al. (2013), little to no change in the relative abundance of prokaryotic community composition in the lower termite *Reticulitermes* under different feeding regimens (filter paper, pine wood slivers, paper and wood) at the end of seven days were observed. Similar to other feeding experiments, the *Reticulitermes* gut profiles under different feeding conditions were compared at the end of the anticipated feeding period. They suggested that the apparent resilience of the prokaryotic community composition is the result of functional redundancy of prokaryotes or the functional independence of this community from lignocellulose digestion (Boucias et al., 2013). Our findings also showed comparable relative abundance of some bacterial populations (Acidobacteria, Bacteroidetes, Fusobacteria, Planctomycetes, Proteobacteria, Synergistetes, Tenericutes and Verrucomicrobia) across all *Mastotermes* gut profiles regardless of time or diets. This suggests that either no change in the bacterial community function was required or that functional changes occurred only at the transcriptional and translational levels which were not detectable via SSU rRNA profiling. While our study supports findings from Boucias et al. (2013) that gut community composition does not reflect the artificial (sugarcane and cotton) feeding treatments at day seven, the compositional shifts in relative abundance of members of Parabasalia, Euryarchaeota, Actinobacteria and Firmicutes in the gut profiles of *Mastotermes* possibly reflect on the failure of the gut microbiota to maintain community stability (**Figure 3.5** and **Figure 3.6**). This is because group size and rate of social interactions are important factors to longevity and vigour of termite colonies (Miramontes and DeSouza, 1996), which are variables not usually considered explicitly in feeding experiments. It is worth noting that in the gut of lower termites, the protist-prokaryotes symbiotic complex plays an important role in lignocellulose degradation and sustaining gut gradients, and any changes in the number of protists may lead to an increase in bacterial populations.

A complementary multivariate PCA analysis comparing the relative abundance of community composition of *Mastotermes* under different feeding conditions showed segregation at days four and seven within each feeding treatment (**Figure 3.7**). We noted that the compositional shifts in *Mastotermes* gut profiles between different time points across all diets were statistically significant (**Figure 3.7**). The compositional shifts observed over a seven day period may be due to declining termite health rather than dietary adaptation. Previous feeding assays that lasted for 21-days for *Nasutitermes* (Miyata et al., 2007) and 30-days for *Coptotermes* (Tanaka et al., 2006) respectively showed gradual mortality of individuals as the trial progressed. *Mastotermes* colonies have successfully been reared for years in laboratory within (Cochrane, personal communication) and outside (Fröhlich et al., 1999; Li et al., 2003) of Australia, in huge enclosed spaces and in large numbers. However, there seems to be a breakdown of the gut microbial community stability when small numbers of *Mastotermes* are placed in smaller enclosed containers over a period of seven days, suggesting that these giant termites are less robust in reduced groups, or that our protocol may have been too restrictive for the survival of this species. In addition, the gradual mortality that was also observed in *Nasutitermes* and *Coptotermes* may simply reflect that although termites are adaptable to changes in diet in the wild, when placed under laboratory conditions, especially in small numbers, it promotes a detrimental physiological response that ultimately affects the gut microbiota. Hence, a precaution should be taken when conducting future feeding experiments and concluding results from such experiments.

Despite stress response as a likely factor for the shift in gut community composition, we observed that there were greater shifts occurring on more compositionally different feedstocks (wood vs. grass and cotton) (**Figure 3.8**). Several reasons that may account for this observation are as follows; (1) shift pattern in wood diet indicates stress response over time; (2) shift pattern between gut profiles on wood and artificial diets indicates dietary response, resulting in the inability of the gut microbiota to respond to the compositional structure of the artificial diets. Although findings from our feeding experiments demonstrated changes in community composition as a dietary response (**Figure 3.8**), supporting our initial hypothesis that predicted membership evenness changes with diet, most of the effects were due to stress. Presumably changes in relative abundances occur over short evolutionary timescales which could account for at least 10,000 years to adjust to new diet and lifestyles (Quercia et al., 2014) fluctuations in population evenness as an adaptation to diet may take longer than seven days with minimum physiological disturbances.

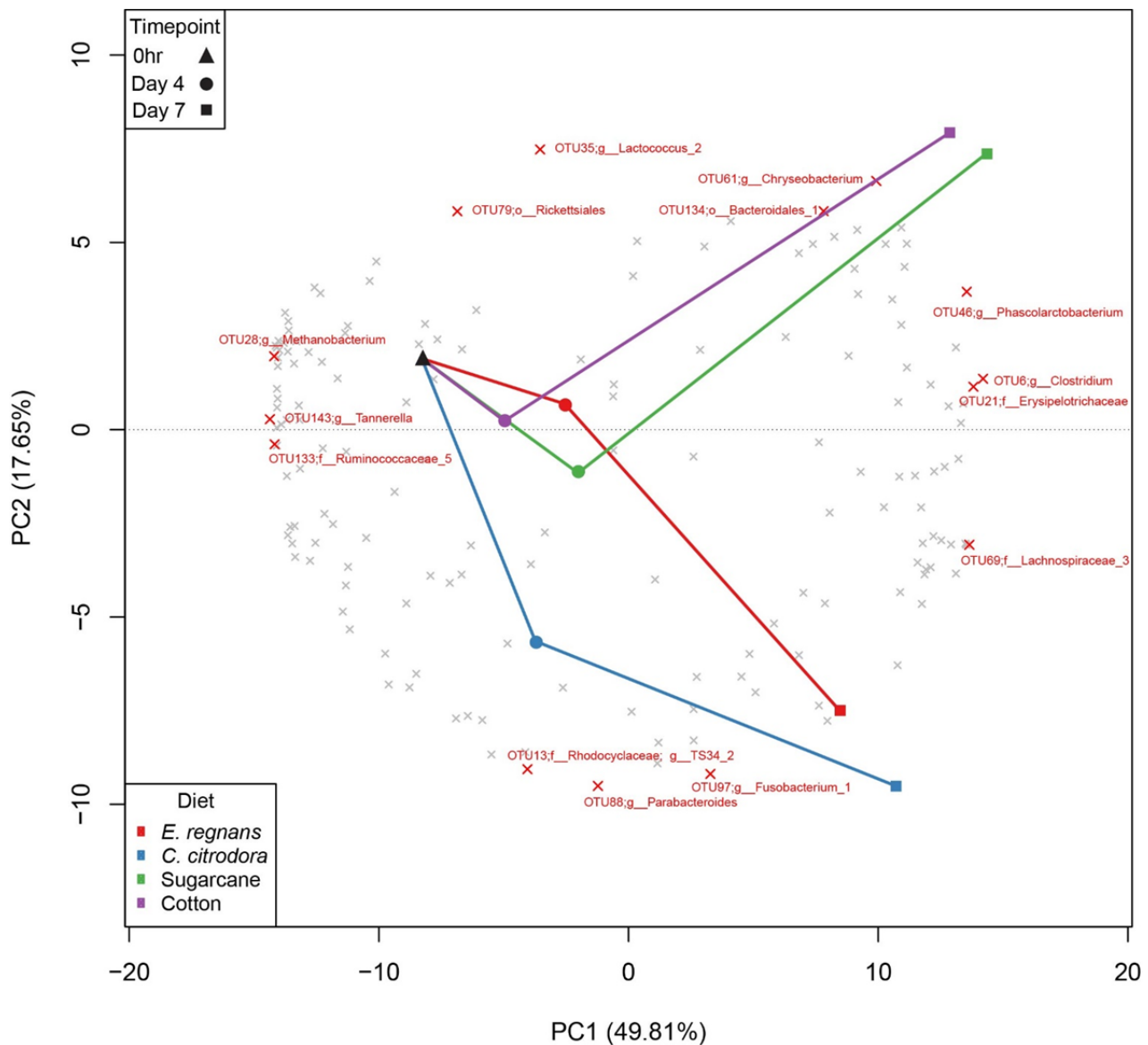


Figure 3.8: Principle component analysis (PCA) plots of microbial profiles obtained from *Mastotermes* gut samples average of biological replicates at 0 hr, Day 4 and Day 7. The colour indicates dietary treatments and shape represents time points. Principal components 1 and 2 explained 49.81% and 17.65% of the variance, respectively.

3.3.3 Identification of active cellulases

Proteomics in combination with zymograms, a functional assay for analysing proteolytic activity (Leber and Balkwill, 1997), were used to determine different lignocellulose feeding impact at the translational level. As this is an optimisation stage of applying proteomics and zymograms to the termite gut, focus was placed on identification of cellulolytic enzymes involved in hydrolysing feedstocks used in feeding trial 2. To generate crude extract for SDS gel electrophoresis, pools of ~20 *Mastotermes* whole guts were homogenised and quantified. The total protein concentration of the obtained crude extract for different feedstocks ranged from 6.2 to 11.8 mg/ml (Appendix B; **Figure S3.2**). Firstly, for the purpose of testing the in-gel substrate zymogram technique, an

experiment was performed on *E. regnans* at time zero to identify cellulase activity. Carboxymethyl cellulase (CMCase) activities were detected through zones of clearing on the zymogram gels in SDS-PAGE under both non-reduced and reduced conditions. Eight clearing zones were observed under non-reduced conditions and only one from reduced conditions (**Figure 3.9**). Due to the nature of protein migration in the SDS-PAGE gel, reduced conditions were chosen for subsequent testing. Protein in non-reduced conditions migrates as a smear and retained activity during electrophoresis which indicates overlapping of clearing regions between different enzymes, making it difficult to interpret the relationship of protein bands to activity zones and isolate individual enzymes contributing to the activity. Separation of protein under non-reduced conditions also resulted in poor resolution on Coomassie-stained gels. By contrast, the zone of clearing (MTZ-1; **Figure 3.10**) detected under reduced conditions corresponded to a band of ~40 kDa (MTC-1 and MTC-2 (replicate); **Figure 3.10**) as guided by the reference Coomassie-stained gel (**Figure 3.10**). MTZ-1 was identified via two approaches; band excised from zymogram gel and reference gel. In total, five samples were trypsin-digested and subjected to mass spectrometry analysis; MTZ-1, MTC-1, MTC-2 and two control *A. niger* from both zymogram (ANZ-1) and reference (ANC-1) Coomassie gels.

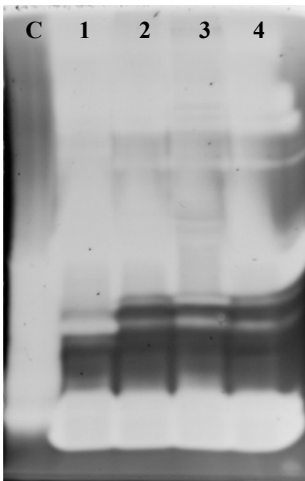


Figure 3.9: Zymograms of crude extract from *M. darwiniensis*. Samples were treated and separated on SDS-PAGE gel under non-reducing conditions. Abbreviations; C: control; 1 to 4: *E. regnans* performed in replicates.

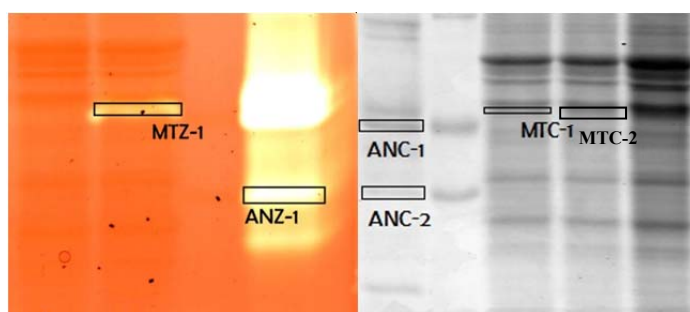


Figure 3.10: Zymogram (left) and corresponding reference gel (right) of crude extract from *M. darwiniensis*. Samples were treated and separated on SDS-PAGE gel under reducing conditions.

All proteins were identifiable by comparison to Swiss-Prot and LudwigNR databases except for ANZ-1 (**Table 3.2**). For bands excised from the zymogram gel, ANZ-1 did not match any cellulose-like proteins while MTZ-1 was identified as a ‘putative glycoside hydrolase (GH) family 7’ from an uncultured symbiotic protist of *M. darwiniensis* (**Table 3.2**). For bands excised from the reference gel, ANC-2 was identified as an ‘endoglucanase A’ from *A. kawachii* (strain NBRC 4308) and both bands MTC-1 and MTC-2 as ‘putative GH family 7’ from an uncultured symbiotic protist of *M. darwiniensis*. The MTC-1 also had a match to ‘beta-1,4-endoglucanase precursor’, a GH9 cellulase originating from *M. darwiniensis*. We also performed peptide searches against the metagenome database (*see* Chapter 4) to identify corresponding contigs. The matching contig ‘contig_87399_1’ was blasted against the NCBI database and had 100% similarity to putative GH7 family (A4UX17), which is the same entry as the Mascot search (**Table 3.2**). This suggests that the metagenome database provides a source to identify novel enzymes that are not publicly available.

Table 3.2: Description of protein hits as determined by mass spectrometry analysis of trypsin digested gel bands from zymogram gel (Z) and Coomassie-stained gel (C).

Band ID	Protein name	CAZy* family	Accession number	Peptides matched	Source
MTZ-1	Putative glycoside hydrolase family7	GH7	A4UX17	49	uncultured symbiotic protist of <i>M. darwiniensis</i>
MTC-1	Putative glycoside hydrolase family7; Beta-1,4-endoglucanase precursor	GH7; GH9	A4UX17 Q81FU4	11 6	uncultured symbiotic protist of <i>M. darwiniensis</i> ; <i>M. darwiniensis</i>
MTC-2	Putative glycoside hydrolase family7	GH7	A4UX17	14	uncultured symbiotic protist of <i>M. darwiniensis</i>
ANZ-1	Keratin	-	e.g. P35527	>400	Various including human
ANC-2	Endoglucanase A	GH12	Q12679	137	<i>Aspergillus kawachii</i> (strain NBRC 4308)

*CAZy – Carbohydrate Active Enzymes

As an alternative approach, overlay zymograms on gradient (8-16%) SDS-PAGE gel was used to examine cellulolytic activity of the gut content of *M. darwiniensis* under non-reducing conditions. In comparison to in-gel zymograms under non-reduced conditions, the overlay zymogram technique allowed a maximum preservation of enzymatic activity while achieving good protein separation for further band identification by mass spectrometry. Comparison of cellulolytic activity between different lignocellulose feeding revealed that there was insignificant alteration in cellulase profiles across the four feedstocks (**Figure 3.11** and **3.12**). The measurement of endoglucanase activity (**Figure 3.11**) confirms that changes in cellulase profiles are very subtle as observed in in-gel zymography (**Figure 3.12**). All samples displayed similar activities (in-gel zymography) and at least five notable enzymes with sizes of ~21 kDa, 33 kDa, 36 kDa, 100 kDa and 120 kDa are responsible for these activities (**Figure 3.12**). The small differences in protein profiles observed suggest that gut function was maintained even though community composition altered. Of the five clearing zones, the two enzymes (sizes ~21 kDa and 36 kDa) that have the most intense activities were MS-21 and MS-36 respectively (**Figure 3.12**). MS-21 and MS-36 were excised from the corresponding reference gel for identification where obtained peptides were affiliated with termite (or insect) and protist proteins according to the SwissProt and LudwigNR databases. However, identified proteins were not associated with cellulase activity. Obtained peptides were also searched against our *Mastotermes* metagenome database and identified 23 contigs that matched the peptide mass spectra, but none were GH-related proteins. This probably indicates that there was low abundance of target enzymes compared to all other proteins present at the excised gel location. To achieve better identification, sample separation by two-dimensional gel electrophoresis can be used to obtain ‘pure’ protein spots. Alternatively, samples can be fractionated using size-exclusion and ion-exchange chromatography. Fractions can then be tested for cellulolytic activity, allowing further sequencing of only relevant fractions.

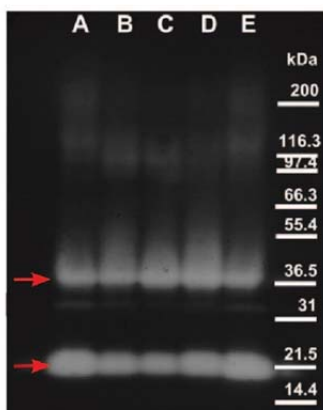


Figure 3.11: Zymography of endoglucanases produced by *M. darwiniensis* gut communities on following diet: A and B – *E. regnans*; C – *C. citrodora*; D - sugarcane mulch; E - cotton. Time points: A – 0 hr; B-E – Day 7.

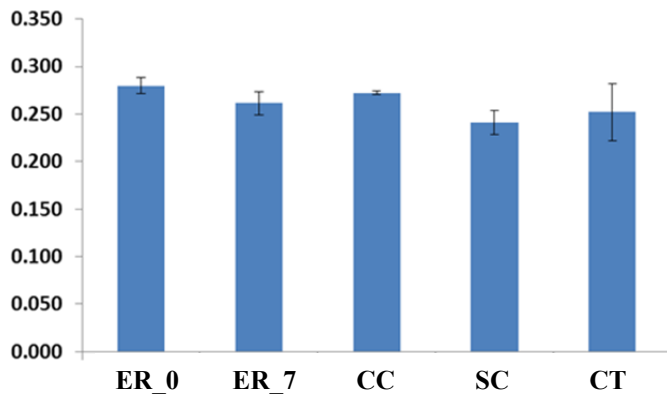


Figure 3.12: Endoglucanase activity (U/mg protein) measured in gut crude extract of termites *M. darwiniensis* subjected to different diet. Values are average from three groups per diet. Error bars represent standard deviation. Abbreviations; ER_0: *E. regnans* at 0 hr; ER_7: *E. regnans* at day 7; CC: *C. citrodora*; SC: sugarcane mulch; CT: cotton.

3.4 Conclusion

In summary, this study provides an insight into the influence of key factors on the survival rate of *Mastotermes* under laboratory conditions and a general assessment of the *Mastotermes* gut microbial community on wood and artificial diets via the use of SSU amplicon sequencing. Overall, our findings support the hypothesis that a seven day feeding period is sufficient to notice the impact of an alternate diet on termite gut microbial communities. The differences in prokaryotic gut composition between wood and alternative-fed *Mastotermes* observed were small suggesting that greater shifts in gut microbiota could be due to stress response to rearing conditions and sources of nutrition as previously described (Husseneder et al., 2009). As some bacterial populations play crucial roles in the survival of termites, selective pressures also ensure that important bacterial phyla (core microbiome) are always present and are contributing factor to the notable community composition shifts. Whether the subtle differences between artificial and wood-fed *Mastotermes* indicates functional redundancy of the prokaryotic community, or that prokaryotic microbiota plays a trivial role in lignocellulosic digestion, or perhaps due to the seven day period, shifts in prokaryotic community composition indicate declining termite health, it will not be known until similar feeding studies are performed (and reproduced) to investigate these communities at the transcriptional and translational level as previously observed in a seven day diet-influenced host and protist symbionts gene expression (Raychoudhury et al., 2013; Sethi and Scharf, 2013). Though our zymogram approach identified protist-derived expressed enzymes, it did not detect any cellulases of bacterial origin, keeping in mind that we are still optimising our protocol, hence requires further testing. These findings contribute new knowledge in understanding of the symbiosis between the termite host and its gut microbiota via symbiont-assisted lignocellulose digestion, providing

important information for subsequent feeding trials and emphasis on the limitations of amplicon sequencing. Hence, this chapter provides the basis to Chapter 4 where shotgun community profiling is compared to amplicon sequencing and also provides insight into the potential roles of prokaryotic communities in lignocellulose digestion.

3.5 Acknowledgements

We thank Micheal Neal and Brian Thistleton for supplying us with *Mastotermes darwiniensis* and Maria Chuvochina for help with setting up feeding trials and protein work. We thank Nicola Angel for Mitag amplicon sequencing and Serene Low for assistance with preparing samples for Illumina sequencing. The study was supported by funding from the Queensland Smart Futures Fund (Future biofuels).

3.6 References

Allison, S.D., and Martiny, J.B. (2008) Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences* **105**: 11512-11519.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Andersen, A.N., and Jacklyn, P. (1993) *Termites of the top end*: CSIRO Publishing.

Angly, F.E., Dennis, P.G., Skarshewski, A., Vanwonderghem, I., Hugenholtz, P., and Tyson, G.W. (2014) CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome* **2**: 11.

Berlanga, M., Paster, B.J., and Guerrero, R. (2009) The taxophysiological paradox: changes in the intestinal microbiota of the xylophagous cockroach *Cryptocercus punctulatus* depending on the physiological state of the host. *International Microbiology* **12**: 227-236.

Boucias, D.G., Cai, Y., Sun, Y., Lietze, V.U., Sen, R., Raychoudhury, R., and Scharf, M.E. (2013) The hindgut lumen prokaryotic microbiota of the termite *Reticulitermes flavipes* and its responses to dietary lignocellulose composition. *Molecular Ecology* **22**: 1836-1853.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**: 248-254.

Brune, A. (2011) Methanogens in the Digestive Tract of Termites. In *(Endo)symbiotic Methanogenic Archaea*. Hackstein, J.H.P. (ed): Springer Berlin / Heidelberg, pp. 81-100.

Brune, A., and Stingl, U. (2006) Prokaryotic symbionts of termite gut flagellates: phylogenetic and metabolic implications of a tripartite symbiosis. *Progress in Molecular and Subcellular Biology* **41**: 39-60.

Butera, G., Ferraro, C., Alonzo, G., Colazza, S., and Quatrini, P. (2015) The gut microbiota of the wood-feeding termite *Reticulitermes lucifugus* (Isoptera; Rhinotermitidae). *Annals of Microbiology*: 1-8.

- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335-336.
- Cardoso, A.M., Cavalcante, J.J., Vieira, R.P., Lima, J.L., Grieco, M.A.B., Clementino, M.M. et al. (2012) Gut bacterial communities in the giant land snail *Achatina fulica* and their modification by sugarcane-based diet. *PloS One* **7**: 149-153.
- Coughlan, M.P., and Moloney, A.P. (1988) Isolation of 1, 4- β -d-glucan 4-glucanohydrolases of *Talaromyces emersonii*. *Methods in Enzymology* **160**: 363-368.
- Darling, A.E., Jospin, G., Lowe, E., Matsen IV, F.A., Bik, H.M., and Eisen, J.A. (2014) PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* **2**: e243.
- Davenport, J. (2012) *Environmental stress and behavioural adaptation*: Springer Science & Business Media.
- Dietrich, C., Köhler, T., and Brune, A. (2014) The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Applied and Environmental Microbiology* **80**: 2261-2269.
- dos Santos Castro, L., Antoniêto, A.C.C., Pedersoli, W.R., Silva-Rocha, R., Persinoti, G.F., and Silva, R.N. (2014) Expression pattern of cellulolytic and xylanolytic genes regulated by transcriptional factors XYR1 and CRE1 are affected by carbon source in *Trichoderma reesei*. *Gene Expression Patterns* **14**: 88-95.
- Erlich, Y., Mitra, P.P., McCombie, W.R., and Hannon, G.J. (2008) Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. *Nature Methods* **5**: 679-682.
- Fröhlich, J., Sass, H., Babenzien, H.-D., Kuhnigk, T., Varma, A., Saxena, S. et al. (1999) Isolation of *Desulfovibrio intestinalis* sp. nov. from the hindgut of the lower termite *Mastotermes darwiniensis*. *Canadian Journal of Microbiology* **45**: 145-152.
- Fuller, C.A., and Postava-Davignon, M. (2014) Termites like it hot and humid: the ability of arboreal tropical termites to mediate their nest environment against ambient conditions. *Ecological Entomology* **39**: 253-262.
- Fuller, C.A., Postava-Davignon, M.A., West, A., and Rosengaus, R.B. (2011) Environmental conditions and their impact on immunocompetence and pathogen susceptibility of the Caribbean termite *Nasutitermes acajutlae*. *Ecological Entomology* **36**: 459-470.
- Geib, S.M., Filley, T.R., Hatcher, P.G., Hoover, K., Carlson, J.E., del Mar Jimenez-Gasco, M. et al. (2008) Lignin degradation in wood-feeding insects. *Proceedings of the National Academy of Sciences* **105**: 12932-12937.
- Gomez-Alvarez, V., Teal, T.K., and Schmidt, T.M. (2009) Systematic artifacts in metagenomes from complex microbial communities. *The ISME Journal* **3**: 1314-1317.
- Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G. et al. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research* **21**: 494-504.

- Harmon-Smith, M., Celia, L., Chertkov, O., Lapidus, A., Copeland, A., Glavina Del Rio, T. et al. (2010) Complete genome sequence of *Sebaldella termitidis* type strain (NCTC 11300(T)). *Standards in Genomic Sciences* **2**: 220-227.
- Hasan, N.A., Young, B.A., Minard-Smith, A.T., Saeed, K., Li, H., Heizer, E.M. et al. (2014) Microbial community profiling of human saliva using shotgun metagenomic sequencing. *PloS One* **9**: e97699.
- Howick, C.D., Creffield, J.W., and Lenz, M. (1975) Field collection and laboratory maintenance of *Mastotermes darwiniensis* Froggatt (Isoptera: Mastotermitidae) for biological assessment studies. *Australian Journal of Entomology* **14**: 155-160.
- Huang, X.-F., Bakker, M.G., Judd, T.M., Reardon, K.F., and Vivanco, J.M. (2013) Variations in diversity and richness of gut bacterial communities of termites (*Reticulitermes flavipes*) fed with grassy and woody plant substrates. *Microbial Ecology* **65**: 531-536.
- Husseneder, C., Berestecky, J., and Grace, J. (2009) Changes in composition of culturable bacteria community in the gut of the Formosan subterranean termite depending on rearing conditions of the host. *Annals of the Entomological Society of America* **102**: 498-507.
- Husseneder, C., Ho, H.-Y., and Blackwell, M. (2010) Comparison of the bacterial symbiont composition of the Formosan subterranean termite from its native and introduced range. *The Open Microbiology Journal* **4**: 53.
- Ikeda-Ohtsubo, W., Desai, M., Stingl, U., and Brune, A. (2007) Phylogenetic diversity of 'Endomicrobia' and their specific affiliation with termite gut flagellates. *Microbiology* **153**: 3458-3465.
- Ives, A.R., and Carpenter, S.R. (2007) Stability and diversity of ecosystems. *Science* **317**: 58-62.
- Knapp, W.R., Holt, D.A., and Lechtenberg, V.L. (1975) Hay preservation and quality improvement by anhydrous ammonia treatment. *Agronomy Journal* **67**: 766-769.
- Knobbe, N., Vogl, J., Pritzkow, W., Panne, U., Fry, H., Lochotzke, H., and Preiss-Weigert, A. (2006) C and N stable isotope variation in urine and milk of cattle depending on the diet. *Analytical and Bioanalytical Chemistry* **386**: 104-108.
- Korb, J., and Fuchs, A. (2006) Termites and Mites - Adaptive behavioural responses to infestation? *Behaviour* **143**: 891-907.
- Köhler, T., Dietrich, C., Scheffrahn, R.H., and Brune, A. (2012) High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Applied and Environmental Microbiology* **78**: 4691-4701.
- Kunin, V., and Hugenholtz, P. (2010) PyroTagger : A fast , accurate pipeline for analysis of rRNA amplicon pyrosequence data. *The Open Journal*: 1-8.
- Kunin, V., Engelbrekton, A., Ochman, H., and Hugenholtz, P. (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* **12**: 118-123.
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

- Lane, D.J. (1991) Nucleic acid techniques in bacterial systematics. In *Nucleic Acid Techniques in Bacterial Systematics*. M., S.E.G. (ed): Wiley, New York, pp. 115-175.
- Leber, T.M., and Balkwill, F.R. (1997) Zymography: A single-step staining method for quantitation of proteolytic activity on substrate gels. *Analytical Biochemistry* **249**: 24-28.
- Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R., and Gordon, J.I. (2008) Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology* **6**: 776-788.
- Li, L., Fröhlich, J., Pfeiffer, P., and König, H. (2003) Termite gut symbiotic archaezoa are becoming living netabolic fossils. *Eukaryotic Cell* **2**: 1091-1098.
- Little, A.E., Robinson, C.J., Peterson, S.B., Raffa, K.F., and Handelsman, J. (2008) Rules of engagement: Interspecies interactions that regulate microbial communities. *Annual Review of Microbiology* **62**: 375-401.
- Luo, C., Tsementzi, D., Kyrpides, N., Read, T., and Konstantinidis, K.T. (2012) Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PloS One* **7**: e30087.
- Lupoi, J.S., Healey, A., Singh, S., Sykes, R., Davis, M., Lee, D.J. et al. (2015) High-throughput prediction of Acacia and eucalypt lignin syringyl/guaiacyl content using FT-Raman spectroscopy and partial least squares modeling. *BioEnergy Research*: 1-11.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A. et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.
- Mikaelyan, A., Dietrich, C., Köhler, T., Poulsen, M., Sillam-Dussès, D., and Brune, A. (2015) Diet is the primary determinant of bacterial community structure in the guts of higher termites. *Molecular Ecology* **24**: 5284–5295.
- Miller, G.L. (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical Chemistry* **31**: 426-428.
- Minkley, N., Fujita, A., Brune, A., and Kirchner, W. (2006) Nest specificity of the bacterial community in termite guts (*Hodotermes mossambicus*). *Insectes Sociaux* **53**: 339-344.
- Miramontes, O., and DeSouza, O. (1996) The nonlinear dynamics of survival and social facilitation in termites. *Journal of Theoretical Biology* **181**: 373-380.
- Miyake, S., Ngugi, D.K., and Stingl, U. (2015) Diet strongly influences the gut microbiota of surgeonfishes. *Molecular Ecology* **24**: 656-672.
- Miyata, R., Noda, N., Tamaki, H., Kinjyo, K., Aoyagi, H., Uchiyama, H., and Tanaka, H. (2007) Influence of feed components on symbiotic bacterial community structure in the gut of the wood-feeding higher termite *Nasutitermes takasagoensis*. *Bioscience, Biotechnology, and Biochemistry* **71**: 1244-1251.
- Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L. et al. (2011) Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**: 970-974.
- Ohkuma, M. (2008) Symbioses of flagellates and prokaryotes in the gut of lower termites. *Trends in Microbiology* **16**: 345-352.

- Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M.H.H., Oksanen, M.J., and Suggests, M. (2007) The vegan package. *Community Ecology Package*: 631-637.
- Quercia, S., Candela, M., Giuliani, C., Turrone, S., Luiselli, D., Rampelli, S. et al. (2014) From lifetime to evolution: timescales of human gut microbiota adaptation. *Frontiers in Microbiology* **5**.
- Quince, C., Lanzén, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M. et al. (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods* **6**: 639-641.
- Ratan, A., Miller, W., Guillory, J., Stinson, J., Seshagiri, S., and Schuster, S.C. (2013) Comparison of sequencing platforms for single nucleotide variant calls in a human sample. *PLoS One* **8**: e55089.
- Raychoudhury, R., Sen, R., Cai, Y., Sun, Y., Lietze, V.U., Boucias, D., and Scharf, M. (2013) Comparative metatranscriptomic signatures of wood and paper feeding in the gut of the termite *Reticulitermes flavipes* (Isoptera: Rhinotermitidae). *Insect Molecular Biology*.
- Robinson, C.J., Bohannan, B.J.M., and Young, V.B. (2010) From structure to function: the ecology of host-associated microbial communities. *Microbiology and Molecular Biology Reviews* **74**: 453-476.
- Robinson, R.K., and Batt, C.A. (1999) *Encyclopedia of Food Microbiology*: Academic press.
- Sabree, Z.L., and Moran, N.A. (2014) Host-specific assemblages typify gut microbial communities of related insect species. *SpringerPlus* **3**: 138.
- Sabree, Z.L., Huang, C.Y., Arakawa, G., Tokuda, G., Lo, N., Watanabe, H., and Moran, N.A. (2012) Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite *Mastotermes darwiniensis*. *Applied Environmental Microbiology* **78**: 204-210.
- Sethi, A., and Scharf, M.E. (2013) Biofuels: fungal, bacterial and insect degraders of lignocellulose. In: *Encyclopedia of Life Science (eLS)*: Chichester: John Wiley and Sons, Ltd.
- Sethi, A., Slack, J., Kovaleva, E.S., Buchman, G.W., and Scharf, M.E. (2013) Lignin-associated metagene expression in a lignocellulose-digesting termite. *Insect Biochemistry and Molecular Biology* **43**: 91-101.
- Smriga, S., Sandin, S.A., and Azam, F. (2010) Abundance, diversity, and activity of microbial assemblages associated with coral reef fish guts and feces. *FEMS Microbiology Ecology* **73**: 31-42.
- Strassert, J.F., Köhler, T., Wienemann, T.H., Ikeda-Ohtsubo, W., Faivre, N., Franckenberg, S. et al. (2012) 'Candidatus Ancillula trichonymphae', a novel lineage of endosymbiotic *Actinobacteria* in termite gut flagellates of the genus *Trichonympha*. *Environmental Microbiology* **14**: 3259-3270.
- Tai, V., James, E.R., Nalepa, C.A., Scheffrahn, R.H., Perlman, S.J., and Keeling, P.J. (2015) The role of host phylogeny varies in shaping microbial diversity in the hindguts of lower termites. *Applied and Environmental Microbiology* **81**: 1059-1070.
- Tanaka, H., Aoyagi, H., Shina, S., Dodo, Y., Yoshimura, T., Nakamura, R., and Uchiyama, H. (2006) Influence of the diet components on the symbiotic microorganisms community in hindgut of *Coptotermes formosanus* Shiraki. *Applied Microbiology and Biotechnology* **71**: 907-917.
- Thompson, C.L., Vier, R., Mikaelyan, A., Wienemann, T., and Brune, A. (2012) 'Candidatus Arthromitus' revised: segmented filamentous bacteria in arthropod guts are members of *Lachnospiraceae*. *Environmental Microbiology* **14**: 1454-1465.

- Tremblay, J., Singh, K., Fern, A., Kirton, E.S., He, S., Woyke, T. et al. (2015) Primer and platform effects on 16S rRNA tag sequencing. *Frontiers in Microbiology* **6**.
- Ward, N.L., Steven, B., Penn, K., Methé, B.A., and Detrich III, W.H. (2009) Characterization of the intestinal microbiota of two Antarctic notothenioid fish species. *Extremophiles* **13**: 679-685.
- Watanabe, H., Takase, A., Tokuda, G., Yamada, A., and Lo, N. (2006) Symbiotic “Archaezoa” of the primitive termite *Mastotermes darwiniensis* still play a role in cellulase production. *Eukaryotic Cell* **5**: 1571-1576.
- Weiser, J., and Hrdy, I. (1962) Pyemotes-mites as parasites of termites. *Zeitschrift für Angewandte Entomologie* **51**: 94-97.
- Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S.A. et al. (2011) Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**: 105-108.
- Xiao, Z., Storms, R., and Tsang, A. (2005) Microplate-based carboxymethylcellulose assay for endoglucanase activity. *Analytical Biochemistry* **342**: 176-178.

Chapter 4 Metagenome-based analysis of gut communities in lower and higher termites

Abstract

Culture-independent small subunit rRNA (SSU) gene amplicon surveys have been widely used to investigate various termite gut habitats. Protocols and approaches used vary between studies with regards to primers and sequencing platforms. These choices may account for the variations observed between different studies of the same species. Despite these choices, limitations of amplicon-based sequencing are associated with primers and the polymerase chain reaction (PCR) introduces biases that may alter microbial diversity estimates. To better understand the effect of primers and PCR biases on termite gut microbial composition, we obtained metagenomes from lower (*Mastotermes* and *Porotermes*) and higher (*Nasutitermes* and *Microcerotermes*) termites and compared the shotgun metagenome-based community profiles to their corresponding amplicon-based profiles previously obtained in Chapters 2 and 3. The community composition was fairly consistent between the metagenome and amplicon profiles, with greater differences noted between lower and higher termite gut profiles. Despite the compositional differences between lower and higher termites, a gene-centric analysis comparing community functional capability between lower and higher termite metagenomes identified commonalities in essential functions to termite biology which includes polymer hydrolysis, nitrogen fixation, hydrogen metabolism and homoacetogenesis. Overrepresentation of hydrolytic enzymes in prokaryotic communities of lower termites support a previous hypothesis of their involvement in lignocellulose degradation. Using a differential-coverage binning method, 179 genome bins were recovered from metagenomes of *Nasutitermes*, *Microcerotermes* and *Porotermes*. Of these, four substantially completed population genomes represent a known termite gut bacterial phylum, Fibrobacteres/TG3, forming the basis for a comparative genomics analysis in Chapter 5.

4.1 Introduction

Culture-independent characterisation of microbial communities in the environment using small subunit (SSU) ribosomal RNA (rRNA) amplicon sequencing has revealed much greater diversity beyond the traditional culture-based microbiological studies. SSU rRNA-based amplicon sequencing provides rapid profiling of microbial communities from the environment, typically through DNA extraction, polymerase chain reaction (PCR) amplification of single genes and identification of sequences to reference databases. However, important limitations of this approach such as PCR and primer biases that favour the amplification of chimeric sequences which leads to overestimation of diversity should be considered. Advances in “next-generation” DNA sequencing technology have resulted in the development of new culture-independent methods that can use genomic information beyond the confines of a single marker gene, most notably shotgun metagenomics (Willner and Hugenholtz, 2013). A major goal of shotgun metagenomic studies, that allows direct analysis of DNA from environmental samples, is to move beyond community profiling of specific microbial marker genes to provide a less biased view of microbial genomic diversity and functionality of an ecosystem (Hugenholtz and Tyson, 2008).

In lower termites, flagellates are cellulolytic specialists and it has been widely thought that bacterial symbionts play a negligible role in lignocellulosic digestion (Mattéotti et al., 2011; Brune, 2014). Recent work has shown possible cellulose digestion involvement of bacterial symbionts in lower termites via (1) genomic identification of genes related to lignocellulose degradation (Boucias et al., 2013; Do et al., 2014; Yuki et al., 2015), (2) antimicrobial treatments which suggest involvement of cellulolytic prokaryotes in metabolism of carbohydrate and phenolic components of lignocellulose (Peterson et al., 2015) and (3) metabolomic profiling of hindgut bacteria through phosphorolysis of cellobiose or cellodextrins (Tokuda et al., 2014). Though these studies have provided insights into the importance of gut prokaryotes in digestive processes, the involvement of these lignocellulose degrading genes in lower termites is still unclear. Furthermore, in Chapter 2, we observed high abundance of termite genera-specific prokaryotic populations of interest in a few of the Australian termites. Here, we performed a gene-centric analysis of five termite gut metagenomes of two lower termite genera (*Mastotermes* and *Porotermes*) and two higher termite genera (*Nasutitermes* and *Microcerotermes*) to determine if system-specific differences exist between gut bacterial microbiota from lower termites and higher termites with similar diets and association of cellulolytic genes. Also, we used differential coverage binning to obtain near and substantially complete bacterial genomes from the termite gut of *Microcerotermes*, *Nasutitermes* and *Porotermes*.

4.2 Methods

4.2.1 Samples and metagenome sequencing

DNA from samples obtained in Chapter 2 and 3 were used for this study, including two lower termites (*Porotermes* and *Mastotermes*) and two higher termites (*Microcerotermes* and *Nasutitermes*). The second colony of *Mastotermes* was collected from Townsville (S19° 18' 40.086", E146° 44' 33.6444"). Briefly, specimens were collected with their nest material and transported to the laboratory in ventilated plastic containers at room temperature to reduce stress to the insects, removed from their nest material within a day of arriving in the laboratory, stored at -20°C and processed as previous described (Chapter 2). The P3 segment gut content of *Mastotermes* workers were collect as described in He et al. (2013) and was separated into two microtubes and processed independently. One microtube of P3 content was passed through a 0.8µm filter to remove larger microbial cells such as protists (filtered P3 content). Genomic DNA (gDNA) was extracted from termite gut using FastDNA® SPIN kit for Soil (MP Biomedicals, Australia) according to manufacturer's instructions.

DNA libraries were prepared from all gDNA for sequencing using a Nextera XT Sample Preparation Kit (Illumina, San Diego, CA, USA). DNA library concentrations were measured using a QuantIT kit (Molecular probes, Carsbad, CA, USA) and equimolar-pooled for sequencing. Samples were paired-end sequenced between a quarter and two fifths of a flowcell lane each on the following platforms with an average fragment size of 320 bp. *Porotermes* was sequenced on an Illumina Miseq (2 x 250 bp) and *Microcerotermes*, *Nasutitermes* and *Mastotermes* were sequenced on an Illumina HiSeq 2000 (2 x 100 bp).

4.2.2 Community profiling

SSU rRNA gene amplicon sequencing of all samples were performed on GS FLX pyrosequencing system (454 Life Sciences, USA) as previously described in Chapter 2. The microbial community composition was determined from the paired-end metagenome data by extracting and classifying SSU rRNA reads using GraftM (<https://github.com/geronimp/graftM>) based on 32 genetic marker genes via hidden Markov model (HMM) search.

4.2.3 Metagenome assembly and annotation

Paired-end reads were overlapped where possible using Seqprep (<https://github.com/jstjohn/SeqPrep>) and quality trimmed using Nsoni (<http://www.vicbioinformatics.com/software.nsoni.shtml>). A *de novo* assembly of the overlapped and quality trimmed reads was generated

using CLC workbench v6 (CLC Bio, Taipei, Taiwan) with kmer size of 63. Open reading frames (ORFs) were identified and annotated using PROKKA (rapid prokaryotic genome annotation) v1.7 using default settings (Seemann, 2014). The summary statistics are shown in Appendix C: **Table S4.1**.

4.2.4 Gene-centric comparative analysis

Gene-centric analysis was performed to reveal relative representation of gene families between metagenomics datasets. Metagenomes of termite gut communities of *Mastotermes*, *Porotermes*, *Nasutitermes*, and *Microcerotermes* were compared based on functional units of Clusters of Orthologous Groups (COG) and KEGG Orthology (KO) assignments or functional categories. The results were mostly consistent across COG and KO. Results were mostly interpreted based on COGs, and in some cases KOs, where COGs were not available. The abundance of each functional unit is the total count of genes belonging to that unit, adjusted (normalised by sequencing depth) by individual population abundance.

4.3 Results and Discussion

4.3.1 Metagenomic sequencing

Bulk DNAs extracted from termite whole gut samples of two *Mastotermes* (*Mastotermes_DW* and *Mastotermes_TV*), one *Porotermes*, one *Microcerotermes* and one *Nasutitermes* for SSU rRNA-based community profiling (Chapter 2 and 3) were used in the present study. A total of 18.8 and 25.3 Gb, of Illumina 2 x 100 bp were sequenced from higher termites *Microcerotermes* (M) and *Nasutitermes* (N). Similarly, 5.9 Gb of Illumina 2 x 250 bp were sequenced from the lower termites *Porotermes* (P) and 3.9 and 6.3 Gb of 2 x 100 bp from *Mastotermes_DW* (MD) and *Mastotermes_TV* (MT) respectively (**Table 4.1** and Appendix C: **Table S4.1**). Sequence datasets from each termite sample (*Microcerotermes*, *Nasutitermes*, *Porotermes*, *Mastotermes_DW* and *Mastotermes_TV*) were independently assembled for gene-centric analysis (Appendix C: **Table S4.1**).

Table 4.1: Summary statistics of metagenomes.

Metagenome	Source	Estimated metagenome size	GC (%)	Total no. of SSU reads detected
<i>Higher termites</i>				
M	<i>Microcerotermes</i> - Chapter 2	18.8 gb	44	46898
N	<i>Nasutitermes</i> - Chapter 2	25.3 gb	44	55787
<i>Lower termites</i>				
P	<i>Porotermes</i> - Chapter 2	5.9 gb	46	9769
MD	<i>Mastotermes</i> (DW) - Chapter 3	3.9 gb	44	5752
MT	<i>Mastotermes</i> (TV) - Townsville	6.3 gb	40	8610

4.3.2 Microbial community composition

Although environmental surveys using SSU rRNA amplicon sequencing (Chapter 2 and 3) has revolutionised the field of microbial ecology, a major caveat of this approach is primer bias and the potential for chimera formation. Hence a primary motivation of this chapter was to use shotgun sequencing to assess community composition because it is not reliant on PCR. Whole gut microbial community composition from three lower and two higher termites was determined by mining the SSU rRNA genes from the unassembled metagenomic datasets using GraftM (<https://github.com/geronimp/graftM>). A majority of the sequences of the profiled gut communities were bacterial, ranging between 76 to 99% in these samples while archaea made up 1 to 23%. The major populations identified using this approach in the lower termite samples belonged to the phyla Bacteroidetes, Firmicutes and Elusimicrobia and in the higher termite, Spirochaetes and Fibrobacteres (**Figure 4.1**).

The SSU rRNA extracted from the five termite gut metagenomes were compared to the previously detailed SSU amplicon profiles (Chapter 2). From a broader taxonomic grouping, the SSU rRNA pyrotag profiles had higher relative abundance of Euryarchaeota as compared to the SSU rRNA metagenome profiles (**Figure 4.1**), most notably in lower termites (particularly *Porotermes*: 20% and 58% of reads respectively). Despite the discrepancy of Euryarchaeota, the relative abundance of the major bacterial taxa between metagenome and pyrotag profiles are comparable, with the exception of *Porotermes*. For *Mastotermes* samples, *Mastotermes_DW* and *Mastotermes_TV*, both the pyrotag and metagenome data showed high abundance of Bacteroidetes and Firmicutes (**Figure 4.1**). For *Microcerotermes* and *Nasutitermes*, Spirochaetes and Fibrobacteres were abundant in both profiles (**Figure 4.1**).

Without taking Euryarchaeota into consideration, Firmicutes and Elusimicrobia were the most abundant populations in both pyrotag and metagenome profiles of *Porotermes*. Other populations such as Spirochaetes, Bacteroidetes and Proteobacteria are detected in low abundance in the pyrotag profile as compared to the metagenome. The differences in relative abundance of some populations are most likely accounted for by PCR primers and amplification bias that has been a concerning issue with amplicon-based sequencing (Berry et al., 2011). It is therefore important to obtain an accurate representation of the microbial community to capture the richness and diversity of the environment. For this reason, shotgun sequencing provides a less biased approach to amplicon sequencing as it does not rely on targeting and amplifying a specific gene and serves as a way to validate discovery of novel organisms in these targeted surveys (Haas et al., 2011; Poretsky et al., 2014). For example, TM7, Tenericutes and Crenarchaeota were undetected or in low abundance in the amplicon data but were present in 0.2-1.0%, 0.1-3.0% and 0.4-6.0% respectively in the shotgun profiles (**Figure 4.1**).

On a finer scale, the top 70 OTUs were extracted from both sequencing platforms (**Figure 4.2**). Similar to the phylum-level analysis, the dominant populations were comparable. For instance in the *Porotermes* specimen, despite the overestimation of *Methanobrevibacter* (58.2%; OTU 2) in the amplicon-based study (Chapter 2), higher proportions of *Methanobrevibacter* (15.5%, OTU 578) than published termite gut profiles (up to 3%; Brauman et al., 2001) were also observed in the metagenome profile (**Figure 4.2**). This supports our initial findings that the archaeal abundance between specimens is more variable than previously appreciated. Similarly, this trend was also observed in several other populations such as a high abundance of class *Endomicrobia* in *Porotermes* (27.5% of pyrotag reads, 22.6% of metagenome), genus *Fibrobacteres-2* (3.3, 3.1% of pyrotag reads; 4.2%, 5.3% of metagenome), family *MIPL1-46* (2.1%, 3.5% of pyrotag reads; 6.1%, 2.6% of metagenome) and genus *Treponema* (53.6%, 22.0% of pyrotag reads, 58.8%, 65.7% of metagenome) in *Microcerotermes* and *Nasutitermes* respectively. The pyrotag and metagenome community profiles were fairly consistent. Although community profiling from metagenomes provides a better estimation of a snapshot of the environmental diversity as it eliminates chimeric OTUs generated from amplicon sequencing, short SSU rRNA sequences extracted from shotgun data results in low resolution taxonomic classification compared to pyrotags (Haas et al., 2011). Longer sequencing length of up to 300 bp can now be obtained to provide an improved taxonomic resolution with advancement in sequencing platforms such as Illumina Miseq (Thomas et al., 2012).

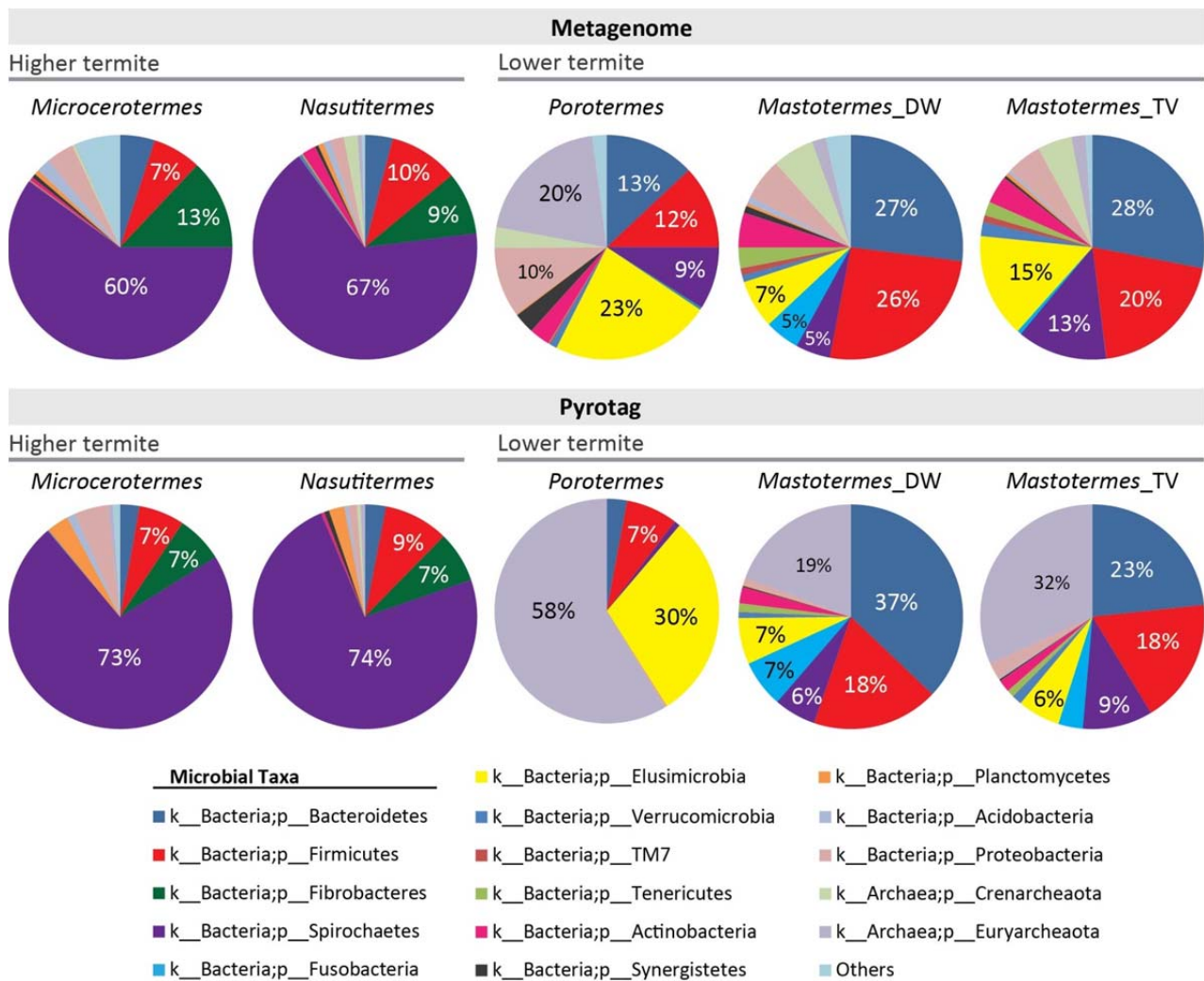


Figure 4.1: Prokaryotic community composition at phylum level by metagenome (top) and pyrotag genes (bottom).



Figure 4.2: Comparison of SSU rRNA gene metagenome-based (metagenome) and amplicon (pyrotag) microbial community profiles based on the top 70 OTUs by total relative abundance across all samples. All population represents >0.2% relative abundance in samples. Samples are displayed in columns accordingly; M; *Microcerotermes*, N; *Nasutitermes*, P; *Porotermes*, MD; *Mastotermes_DW*, MT; *Mastotermes_TV*. OTUs are grouped by phyla and ordered by taxonomy (lowest to highest) and displayed in rows.

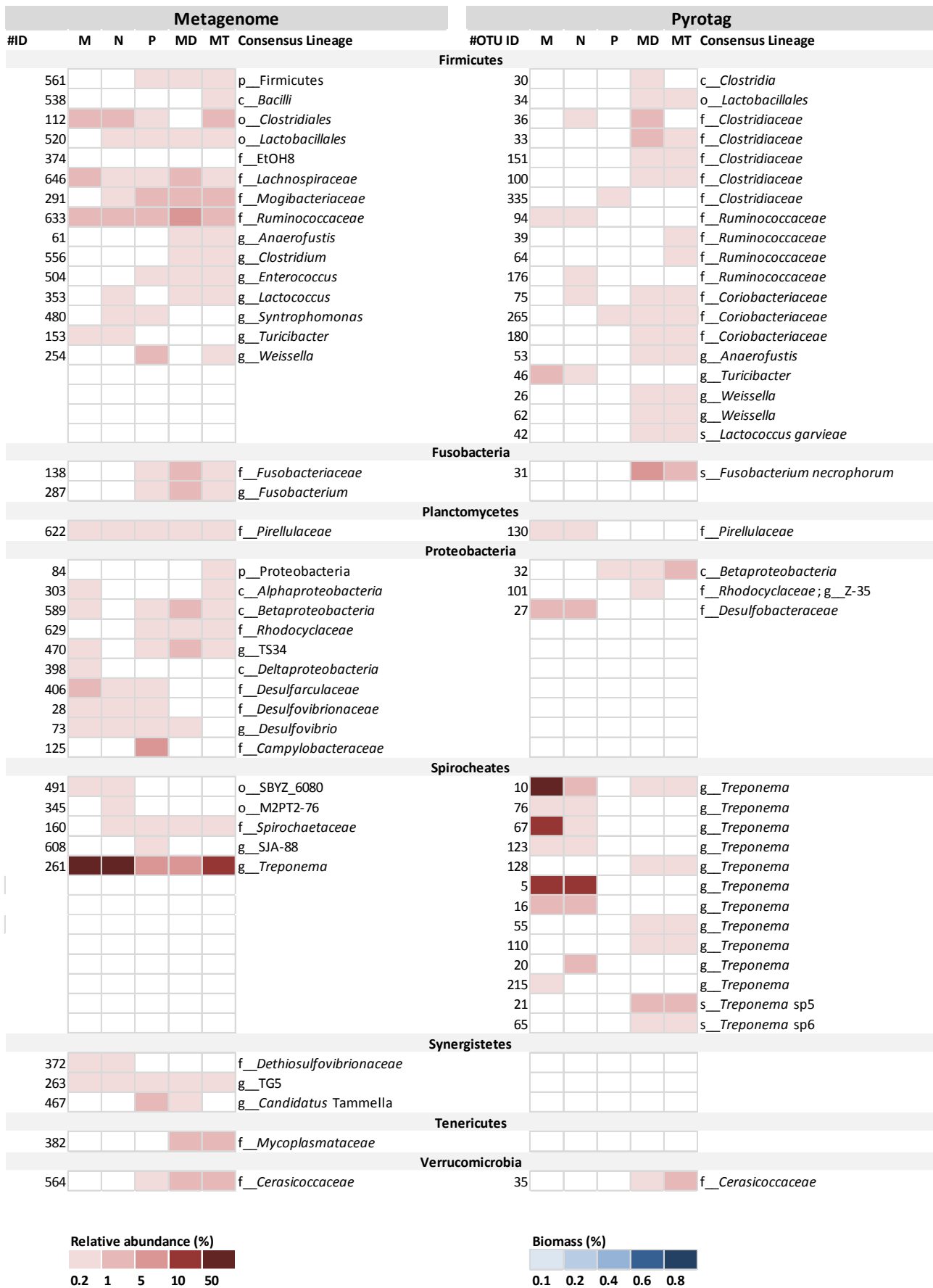


Figure 4.2: Continued.

4.3.3 Gene-centric analysis of community metabolism

Metagenomic analyses were performed to identify functional similarities and differences between lower and higher termite gut microbiota. Metagenome data were generated with Illumina shotgun sequencing. The summary statistics of metagenomic sequencing, assembly and annotation are listed in Appendix C: **Table S4.1**. Gene-centric analysis was conducted to compare the metagenomes based on functional categories of Clusters of Orthologous Groups (COG) and KEGG Orthology (KO) functional units. The higher termite metagenomes shared greater similarities in functional profiles to each other than to the lower termites. Consistencies among the lower termite metagenomes were observed (**Figure 4.3**). There are also functional differences between lower and higher termite gut microbiota, particularly cell motility and associated chemotaxis which were overrepresented in higher termites.

4.3.4 Functional profiling of the DNA metagenomes

Functional profiles were determined by classification of predicted genes based on the COG database. Of the annotated protein coding sequences, 21.8%, 14.6%, 17.8%, 12.9% and 8.4% from *Microcerotermes*, *Nasutitermes*, *Porotermes*, *Mastotermes_DW* and *Mastotermes_TV* had COG functional prediction respectively classified into the 25 different COG categories (**Figure 4.3** and Appendix C: **Table S4.1**). From this analysis, there is an overrepresentation of genes in six categories in all five metagenomes, including translation, ribosomal structure and biogenesis (category J, 8.5-13.7%), amino acid transport and metabolism (E, 8.1-10.2%), replication, recombination and repair (L, 7.8-9.2%), energy production and conversion (C, 5.0-8.1%), carbohydrate transport and metabolism (G, 5.5-6.5%) and cell wall/membrane/envelope biogenesis (M, 5.0-5.5%) (**Figure 4.3**). These results are likely reflective of the biology of the prokaryotic community in the termite gut ecosystem. The enrichment of genes required for cellular processes and metabolism suggest that they are capable of acquiring their own nutritional needs. In higher termites, gut prokaryotic communities are dependent on the metabolism of carbohydrates to perform these cellular processes (Warnecke et al., 2007; He et al., 2013). The moderately abundant genes for COG category M (cell wall/membrane/envelope biogenesis) possibly reflect the need to attach to plant cell walls (Suen et al., 2011). The enrichment of these genes in lower termites suggest that the gut prokaryotic communities, although mostly present as ecto- and endosymbionts of flagellated protists (Brune, 2014), are also capable of these essential functions. It was previously suggested that in honey bees the transportation of sugars from the environment is assisted via the components of multiple phosphotransferase systems which represent the majority of the functions within COG category G (carbohydrate transport and metabolism) (Engel et al., 2012). These sugar

transporters are also detected in our metagenomes (Appendix C: **Table S4.2**), which could explain the abundance of glycoside hydrolases as an adaptation to plant cell wall digestion. Moreover, the general eukaryotic functional categories such as categories Z (cytoskeleton), W (extracellular structures), A (RNA processing and modification), and B (chromatin structure and dynamics) are absent from our metagenomic dataset consistent with a previous study (Do et al., 2014).

Despite the functional commonalities that exist between lower and higher termite gut metagenomes, two COG categories are more abundant in higher than lower termites; category N (cell motility) and T (signal transduction mechanisms). The composition of these categories is in accordance with the termite gut structure and the free-living nature of the bulk of bacteria in higher termites (Appendix C: **Table S4.3** and **Table S4.4**). The steep physiochemical gradients that consist of an anoxic central and a microoxic periphery within the termite gut are known reservoirs of motile bacteria, unlike mammalian rumen that harbours mostly non-motile bacteria (Beckwith and Light, 1927; Brune and Friedrich, 2000). In contrast, bacteria in lower termites are either associated with the flagellates, attached to the gut wall or free-living. Those bacteria that exist as ectosymbionts of flagellated protists have been reported to possess flagella as they are involved in symbiotic motility by propelling their eukaryotic hosts (König et al., 2006; Hongoh et al., 2007). Low abundance of motility genes in the prokaryotic community in lower termites reflects their overall abundance as compared to those in higher termites. The high abundance of cell motility and chemotaxis genes in higher termites (*Microcerotermes* and *Nasutitermes*) to lower termites may also reflect the differences in community composition, specifically highly motile populations such as the Spirochaetes (He et al., 2013). From both the metagenome and pyrotag profiles (**Figure 4.1** and **Figure 4.2**), the Spirochaetes are more abundant in both higher termites (60-74%) as compared to lower termites (1-13%) consistent with previous findings (He et al., 2013).

4.3.5 Plant polysaccharide degradation enzymes

Plant cell walls are mainly composed of lignocellulose that contains 40-55% of cellulose encrusted with 20-40% hemicellulose and 10-20% lignin. Both the host and gut symbionts of termites contribute to lignocellulose degradation but for the scope of this chapter, we only focus on the gut prokaryotic symbionts. The predicted genes encoding glycoside hydrolases (GHs) from the metagenomes were classified into 88 families based on the Carbohydrate-Active enZYmes (CAZy) database (Lombard et al., 2014) (Appendix C: **Table S4.1**). Cellulose and hemicellulose are abundant polysaccharides while lignin is an insoluble complex polymer in plant cell walls. In order to access cellulose and hemicellulose, lignin has to be removed or at least modified (Zabel and Morrell, 2012). However, consistent with previous efforts (Todaka et al., 2007; Warnecke et al.,

2007; He et al., 2013; Do et al., 2014), we did not detect any lignin-degrading enzymes from symbionts of both lower and higher termites. Unlike the fungus-growing termite family *Macrotermitinae* that benefits from lignin digestion by their fungal symbiont, other termites must overcome this barrier through alternative means (Brune, 2014). Recent studies have shown small alterations in the overall lignin structure, in both lower and higher termites, (Hopkins et al., 1998; Hyodo et al., 1999; Katsumata et al., 2007; Geib et al., 2008; Li et al., 2011) and provided evidence of potential expressed lignase genes in the gut of the lower termite *Reticulitermes* (Tartar et al., 2009; Sethi et al., 2012). Even so, lignin is a major end product found in termite faeces, at least in wood-feeding termites (Brune, 2014). It has been suggested that any modification to lignin structure would increase the accessibility of polysaccharides to enzymes (Raychoudhury et al., 2013; Brune, 2014). Along with enzymatic digestion, mechanical breakdown through the hosts mandibles and gizzard also contribute to a decrease in the mass of lignin hence assisting the initial digestion processes (Hyodo et al., 1999).

It is known that in higher termites, prokaryotic communities are the major source of glycoside hydrolases (GHs) for depolymerisation of recalcitrant plant fibre due to the absence of flagellates. In the lower termites, flagellates produce the bulk of GHs in the hindgut paunch and prokaryotic symbionts are thought to not play a significant role in digestion (Mattéotti et al., 2011; Brune, 2014). All metagenomes included in this study had an abundance of genes encoding bacterial GHs for cellulose and hemicellulose degradation (Appendix C: **Table S4.2**). The number of genes encoding GHs in the higher termite metagenomes are five-fold higher than in the lower termites which corresponds to the SSU rRNA community composition (Appendix C: **Table S4.1**). Cellulases and hemicellulases represent 55-65% of total genes encoding GHs in both higher and lower termites. These genes encoding GHs from prokaryotic symbionts of two lower termite genera, *Mastotermes* and *Porotermes*, support potential involvement of prokaryotes in lignocellulose digestion, as recently reported in other lower termites (Mattéotti et al., 2012; Do et al., 2014; Tokuda et al., 2014; Peterson et al., 2015; Yuki et al., 2015).

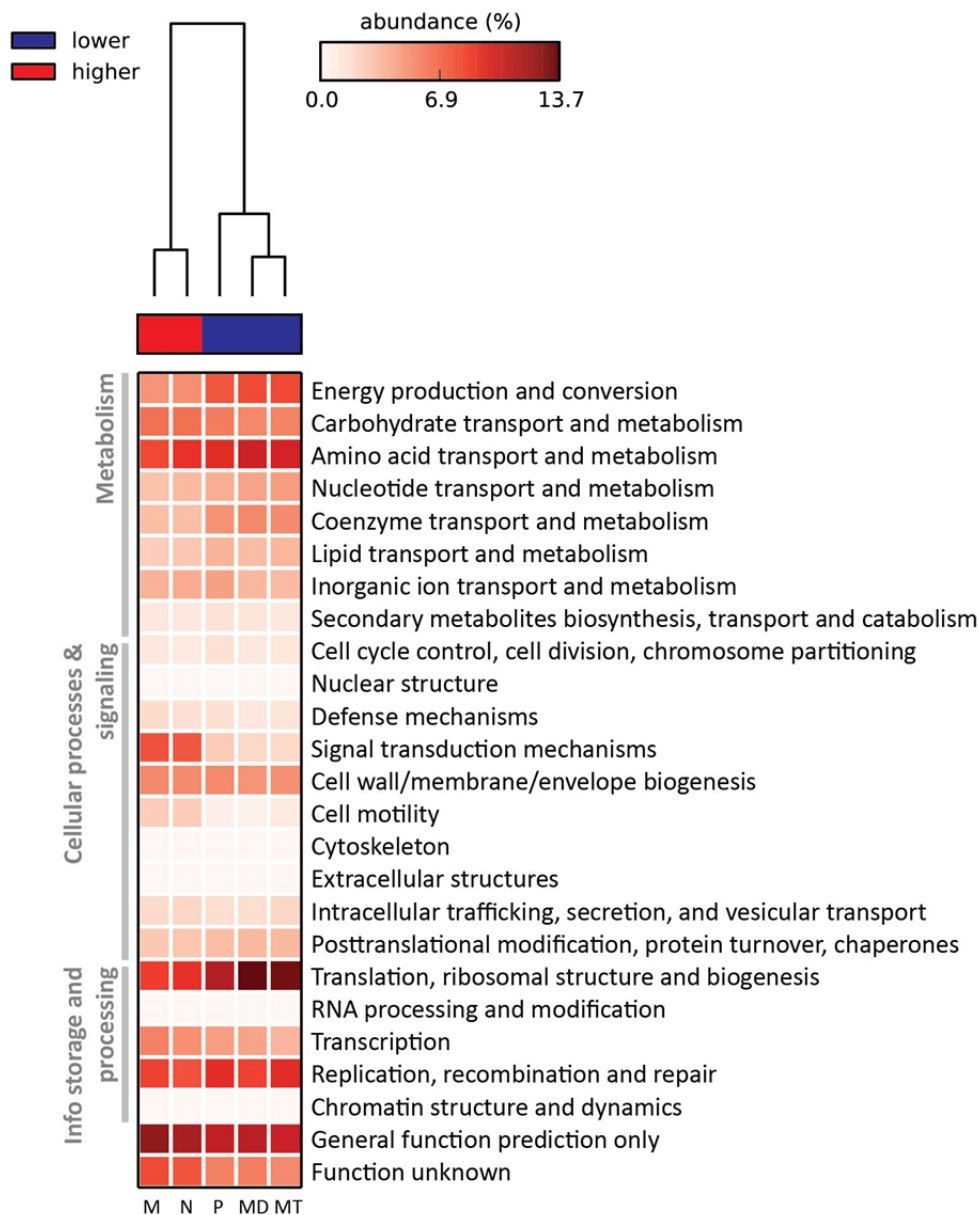


Figure 4.3: Relative abundance of COG functional categories (% of total coding sequences) across the investigated metagenomes; M: *Microcerotermes*, N: *Nasutitermes*, P: *Porotermes*, MD: *Mastotermes_DW*, MT: *Mastotermes_TV*. Metagenomes are ordered by similarity of relative patterns.

From the broad functional categories of GHs, all metagenomes had higher relative abundances of hemicellulases as compared to cellulases, one-fold higher in higher termites and four-fold higher in lower termites (**Figure 4.4** and Appendix C: **Table S4.2**). Cellulases and hemicellulases are major enzymes in plant cell wall degradation that are responsible for hydrolysis of cellulose chains and the side chains of hemicellulose (Breznak and Brune, 1994; Scharf, 2015). The removal of hemicellulose is considered a pre-requisite step into gaining access to cellulose (Scharf and Tartar, 2008; Sethi et al., 2012; Scharf, 2015). The polysaccharide matrix is composed of three branched backbones, xylan, xyloglucan and mannan. Degradation of hemicellulose is achieved through action of multiple hemicellulases targeting these main backbones. Hemicellulases are relatively high in

both the higher (35.9-39.4%) and lower (45.4-49.5%) termite metagenomes. Eighteen GH families with known hemicellulase activities were identified in all five metagenomes (**Figure 4.5**, Appendix D: **Table S4.2** and **Table S4.3**). Twelve of these (GH2, GH10, GH26, GH27, GH28, GH35, GH39, GH42, GH43, GH51, GH53 and GH67) contain catalytic domains of endo-xylanases, endo-1,4- β -xylanases, xyloglucanases, xyloglycosyltransferases, β -mannanases, galacturonases, arabinosases, xylosidases and xylan α -1,2-glucuronosidases that were previously reported in the gut of other termite genera including *Reticulitermes* (Tartar et al., 2009), *Coptotermes* (Do et al., 2014) and *Amitermes* (He et al., 2013). Five of the 18 hemicellulase-containing GHs identified in this study are associated with prokaryotic-derived debranching enzymes that cleave hemicellulose side chains, detected in the gut of fungus-growing termites, *Pseudacanthotermes militaris* and *Macrotermes natalensis* (Bastien et al., 2013; Poulsen et al., 2014). The high relative abundance of hemicellulases in prokaryotic communities in lower termites suggests that they may play an assisting role, to that of protists, to achieve hemicellulose removal.

The cellulases identified in the gut metagenomes of the higher termite genera *Microcerotermes* and *Nasutitermes* represent 28.3% and 28.7% of the overall GHs respectively. In the lower termites *Porotermes*, *Mastotermes_DW* and *Mastotermes_TV*, 10.0%, 10.2% and 9.6% of the GHs were cellulases respectively. Cellulases can be classified into three main classes which include endocellulases (β -1, 4-endoglucanases/cellulases), exoglucanases/cellobiohydrolases and β -glucosidase (Breznak and Brune, 1994; Scharf, 2015). These enzymes work synergistically and sequentially to completely digest cellulose (Breznak and Brune, 1994; Scharf, 2015). Generally, the endoglucanases are responsible for cleaving the main cellulose chains to release polysaccharide units, making it accessible to exoglucanases and β -glucosidase. Exoglucanases hydrolyse the non-reducing ends of side chains to further release cellobiose (a glucose dimer). Cellobiose is then hydrolysed to glucose by glucosidase or cleaved by cellobiose phosphorylase. Within the three main classes, there are differences in abundances of cellulases between lower and higher termites. For instance, two endoglucanases GH5 and GH9 are found in all metagenomes, four-fold higher in relative abundance in higher termites (11.5-11.8%) compared to lower termites (1.5-2.5%) (Appendix C: **Figure S4.1** and **Table S4.2**). This is consistent with previous reports of high abundance of cellulases from gut bacterial symbionts in the higher termite genus, *Nasutitermes*, which are the key players in cellulose degradation (Tokuda and Watanabe, 2007, Warnecke et al., 2007; He et al., 2013). Tokuda and Watanabe (2007) reported that the hindgut of *Nasutitermes* sp. contained several cellulases responsible for up to 59% of cellulase activity against crystalline cellulose in comparison to the midgut. They also showed that termites under antibiotic treatment had a significant decrease in cellulase activity suggesting cellulase enzymes in the hindgut of

Nasutitermes sp. were derived from symbiotic bacteria (Tokuda et al., 2007), further supporting the higher abundance of prokaryotic cellulases in the *Microcerotermes* and *Nasutitermes* metagenomes when compared to the lower termites. Relatively low abundances of GH5 and GH9 suggest that prokaryotic symbionts in lower termites may be peripherally involved in cellulose degradation. Do et al (2014) also reported relatively low GH5 and GH9 in the free-living bacterial community of a lower termite *Coptotermes gestroi*. In addition, at least one or more genes encoding β -glucosidases are identified in all gut metagenomes that are responsible for cleaving cellobiose and cellodextrins (polymer) to glucose. However, we did not find any exoglucanases in our metagenome datasets which was also undetected in *C. gestroi* (Do et al., 2014). It has been reported that in the lower termite *Reticulitermes flavipes*, exoglucanase activity is associated with protist-derived GH7 (Zhou et al., 2007). In the *Mastotermes* proteome in Chapter 3, which corresponds to the *Mastotermes_DW* metagenome, we noted protist-derived GH7 (A4UX17) expression and as expected, no GH7 gene was detected in the prokaryotic-annotated metagenomic data. However, peptide searches against the *Mastotermes_DW* metagenome identified that A4UX17 had 100% similarity to 'contig_87399_1' which confirms the presence of GH7 in the unannotated metagenomic data (see Chapter 3; **Table 3.2**). This supports the inference that protist symbionts are primarily responsible for exoglucanase activity in the guts of lower termites. For higher termites, no exoglucanase has been identified to date, as it may not be a requirement for cellulose digestion in the higher termite gut ecosystem and that exoglucanases are rare in bacterial cellulase systems (Breznak and Brune, 1994). Cellobiose or cellodextrin phosphorylases (GH84 and GH94; Appendix C: **Table S4.2**) were also detected in both the higher and lower termite metagenomes. It was proposed that bacterial cellobiose or cellodextrin phosphorylases in higher termites play an important role in the hydrolysis of wood polysaccharides by catalysing oligosaccharides in the bacterial cytoplasm (Warnecke et al., 2007). Interestingly, the presence of bacterial cellobiose or cellodextrin phosphorylases in lower termite may play similar role to those in higher termite as recently demonstrated in a metabolomics profiling of ¹³C-labelled cellulose digestion (Tokuda et al., 2014). The proposition that *Treponema* is involved in cellulolysis in higher termite (Warnecke et al., 2007) may also be applicable to lower termite (Tokuda et al., 2014). The abundance of GH94 in *Nasutitermes* (2.2%), *Microcerotermes* (2.2%) and *Porotermes* (0.2%) (Appendix C: **Table S4.2**) correspond to the community composition of *Treponema* (**Figure 4.2**), ~50%, ~50% and 5%, in these metagenomes respectively, supporting its relative roles in cellulose digestion in both lower and higher termites.

In the metagenomes, chitinases are present in low abundance (2.1-4.3%) in all metagenomes. Chitin is a major component of the insect exoskeleton (Shen and Jacobs-Lorena, 1997). In insects,

chitinases are secreted to facilitate in moulting (shedding part of its body) by cleaving chitin in the exoskeleton into smaller oligosaccharides (Shen and Jacobs-Lorena, 1997). Chitin is also an important component of the insect peritrophic matrix (PM), a membrane that surrounds food bolus in the midgut of the insect (Tellam, 1996; Shen and Jacobs-Lorena, 1997). It is known that bacteria-derived chitinases in insect guts are responsible for maintaining the physical properties of PM and that the chitinous sheath of PM provides bacteria with a source of nitrogen and carbon for growth and development (Shen and Jacobs-Lorena, 1997; Indiragandhi et al., 2007). Chitinases may also be involved in digestion of cell wall of fungal hyphae ingested in wood particles and dead nest mate bodies (Hongoh, personal communication). This suggests that the chitinases detected in the guts of lower and higher termites may have similar functions.

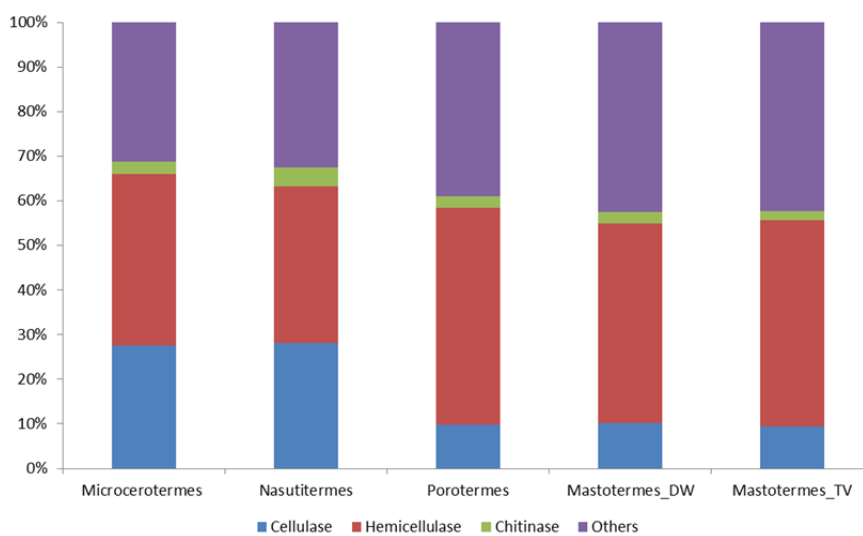


Figure 4.4: Distribution of genes encoding GH family proteins from the termite gut.

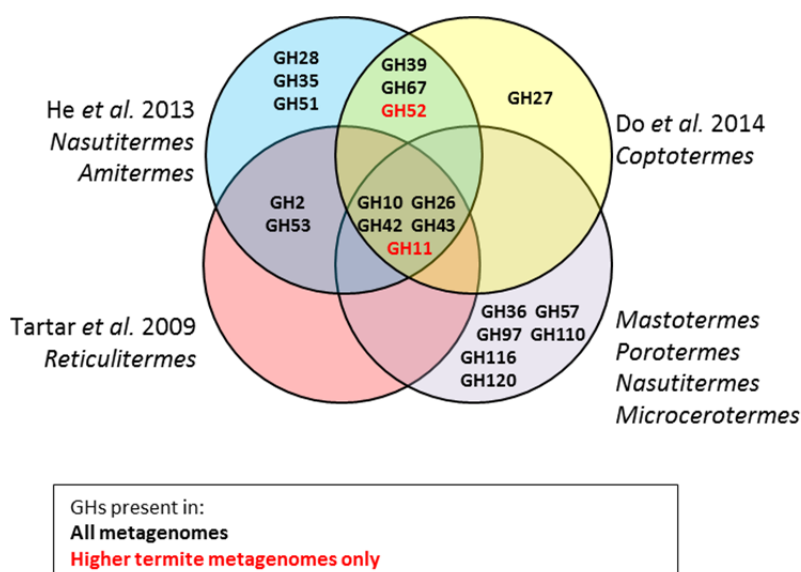


Figure 4.5: Distribution of GH families with known hemicellulase activity in this study and other termite gut.

4.3.6 Other major functions in the termite gut ecosystem

From our initial COG and KO functional categories analyses, we further investigated the enriched genes contributing to the carbohydrate and energy metabolism categories. Genes encoding proteins for the two major metabolic pathways of polymer hydrolysis, glycolysis and pentose phosphate pathways, were relatively abundant in both lower and higher termites (Appendix C: **Table S4.3** and **Table S4.4**), indicating the importance of polysaccharide metabolism in gut ecosystems. Although sugar transporters are more abundant in higher than lower termites, probably due to community composition differences, presence of these transporters in lower termites indicates the potential ability of the prokaryotic community in polymer hydrolysis of imported monosaccharides such as glucose and xylose.

Bacteria are known key players of nitrogen fixation that is an important function in the termite guts as they thrive on diets that are nitrogen poor (Benemann, 1973; Breznak et al., 1973; Brune, 2014). As previously detected by Warnecke et al. (2007) and He et al. (2013), bacteria in the higher termite genera, *Nasutitermes* and *Amitermes*, have rich arrays of nitrogen fixation genes encoding nitrogenase components which were also observed in the higher termite *Microcerotermes* and *Nasutitermes* metagenomes. We also found that the lower termite prokaryotic symbionts have large numbers of nitrogenase genes relative to total protein coding genes (Appendix C: **Table S4.3**). Some bacterial symbionts of termite flagellates have been shown to play essential roles in nitrogen fixation (Noda et al., 1999; Ohkuma et al., 1999; Hongoh et al., 2008; Hongoh, 2011; Brune, 2014).

Hydrogen is an important product generated from the fermentation of cellulose and xylan to acetate via hydrogenase activities. In the higher termite metagenomes, the iron-only hydrogenase encoding genes are abundant in both the *Microcerotermes* and *Nasutitermes* as previously noted (Warnecke et al., 2007; He et al., 2013). Although protists are known to be the main producer of hydrogen through fermentation, abundance of the iron-only hydrogenase genes in the lower termite metagenomes suggest that the prokaryotic communities may also be involved in hydrogen generation (Appendix C: **Table S4.3**). This is in agreement with previous reports that identified homologues of iron-only hydrogenases in both lower and higher termites most closely related to spirochaetes (Leber and Balkwill, 1997; Fröhlich et al., 1999; Warnecke et al., 2007; Brune, 2014). In addition, low abundance of NiFe hydrogenases were detected in all the metagenomic datasets as previously reported (Warnecke et al., 2007).

Another major function that the termite gut symbionts are responsible for is homoacetogenesis (Wood-Ljungdahl pathway). Homoacetogenesis involves the reduction of CO₂ to acetate which is a major H₂ sink process (Ljungdahl, 1986) and ultimately acetate is used by the termite host as a

carbon and energy source (Odelson and Breznak, 1983). The importance of homoacetogenesis in both lower and higher termite gut ecosystems is highlighted by the abundance of genes encoding the Wood-Ljungdahl pathway in all metagenomes (Appendix C: **Table S4.3** and **Table S4.4**). Consistent with previous findings, genes encoding NADPH-dependent formate dehydrogenases required for catalysing the methyl group acetate were present in the lower termite metagenomes (Zhang et al., 2011) but lacking from the higher termite (Warnecke et al., 2007; He et al., 2013). It was hypothesised that an alternative pathway to generate formate was used, likely via the pyruvate formate lyases which were present in all metagenomes (Warnecke et al., 2007). Homologs of cysteine and selenocysteine variants of hydrogenase-linked formate dehydrogenases (FDH(H)) (alternative to NADPH-dependent formate dehydrogenase) that use H₂ instead of NADPH were also present in the higher and lower termite metagenomes (He et al., 2013). Additionally, a key enzyme responsible for homoacetogenesis, formyl tetrahydrofolate synthase (FTHFS) was identified in high abundance across all metagenomes supporting the significance of this metabolic function in the termite gut (Appendix C: **Table 4.4**) (Warnecke et al., 2007; Zhang et al., 2011).

4.3.7 Population genome binning

Gene-centric analysis, while useful for assessing commonalities and differences in overall community function between ecosystems, generally provides limited insight on the populations responsible for those functions (Kunin et al., 2008; Brulc et al., 2009). The approach also tends to overlook potentially important functions that maybe underrepresented and do not provide direct association of the specific populations involved in metabolic pathways of overrepresented genes (Brulc et al., 2009). Hence, population genome binning (assigning sequences into defined taxonomic groups) allows associating metabolic data to organisms that are likely responsible for the processes. It is one of the goals of this chapter to recover population genomes of interest from the metagenomes obtained from selected lower and higher termites identified in Chapter 2. In the initial survey, we identified higher relative abundance of Fibrobacteres from gut microbiota of wood-feeding *Microcerotermes* and *Nasutitermes* which is an underrepresented bacterial phylum but important in the termite gut ecosystem. We also noted high abundance of Elusimicrobia (~65% of bacterial reads; see Chapter 2: **Figure 2.2**) in the lower termite, *Porotermes*. For this reason, we attempted to recover population genomes from co-assemblies of the gut metagenomes of *Microcerotermes* and *Nasutitermes* using differential coverage binning (Albertsen et al., 2013; Imelfort et al., 2014) and *Porotermes* metagenome using distribution-based binning (DBB v1.0.1; <https://github.com/dparks1134/DBB>). A total of 179 population genomes were recovered from the three datasets (Appendix C: **Table S4.1**), and of these, only three from *Microcerotermes*, one from *Nasutitermes*, and one from *Porotermes* were substantially complete (>60% completeness and

<10% contamination) according to CheckM estimates and classification (Parks et al., 2015). The four substantially complete population genomes from *Microcerotermes* and *Nasutitermes* were affiliated with Fibrobacteres and the TG3 candidate phylum, and one from *Porotermes* was affiliated with the order *Lactobacillales*. Dominant bacterial populations such as Spirochaetes and Elusimicrobia in the higher termites and *Porotermes* metagenomes were not binned. This could be due to host DNA contamination, as the whole gut was used instead of lumen contents, which hampers downstream processes such as assembling the reads. Whole gut was initially used to obtain the gut microbiota attached to the gut wall and in the gut lumen. We therefore tested the effect of three different gut specimen parts (whole gut, P3 content and filtered P3 content) from a subset of *Mastotermes* from Chapter 3 on the resulting metagenomic data. The whole gut sample resulted in ~40% of the eukaryotic (insect host, protists and other eukaryota) sequences in the metagenome datasets (Appendix C: **Figure S4.2**). The P3 content and filtered P3 content metagenomes have two-fold lesser eukaryotic sequences (27.3% and 11.4% respectively) as compared to the whole gut. Eukaryotic genomes in metagenomic datasets are often considered as contaminants and are not well characterised due to tandem repeats and polymorphisms (Gilbert and Dupont, 2011; Zhou et al., 2014), which further complicates the quality of the final assembly. Moreover, the recovery of population genomes using differential coverage will be greatly impeded by the quality of the metagenomic assembly (Imelfort et al., 2014). Therefore caution should be exercised when selecting termite gut samples for subsequent shotgun sequencing and preceding processes (e.g. DNA extraction) should be tailored based on the specific goals of analysis.

4.4 Conclusion

This chapter provides a new view into the complex prokaryotic communities in the termite gut. The community composition obtained from the metagenome was comparatively consistent with amplicon profiles, with divergences between the lower and higher termites. The relatively high proportion of archaea in the metagenome community profile of *Porotermes* confirmed the archaeal abundance in the initial amplicon profile. Despite the well-defined dominant role that protists play in the gut of lower termites, the gene-centric analysis revealed that prokaryotic symbionts have genes for key functionality (i.e. nitrogen fixation, hydrogen metabolism and homoacetogenesis), similar to those in higher termites. The abundance of GH families across both lower and higher termites likely contribute unique degradation potential that is shared between these metagenomes. Our findings support the ability of the prokaryotic communities to degrade lignocellulose in the gut of lower termites suggesting that with the evolutionary loss of cellulolytic protists in higher termites, bacterial populations dominate to take over the major roles. The absence of formate

dehydrogenases anomaly previously reported appears to be common across the higher termites and would be interesting to determine if it is a trait in other higher termites. Findings in this chapter also provides foundation for future studies targeting specific roles of important bacterial populations in the gut of both lower and higher termites via population genome binning or metatranscriptomics approaches. We were particularly interested in the poorly represented bacterial phylum Fibrobacteres that were abundant in the gut profiles of wood-feeding higher termites than grass-feeding species (Chapter 2). We therefore conducted a comparative genomic analysis of the phylum Fibrobacteres using the substantially completed Fibrobacteres genomes recovered from *Nasutitermes* and *Microcerotermes*, together with six other Fibrobacteres genomes from other anaerobic habitats (Chapter 5).

4.5 Acknowledgements

We thank Serene Low for help with preparing samples for Illumina sequencing and Donovan Parks for helpful discussions.

4.6 References

- Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., and Nielsen, P.H. (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology* **31**: 533-538.
- Bastien, G., Arnal, G., Bozonnet, S., Laguerre, S., Ferreira, F., Fauré, R. et al. (2013) Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics. *Biotechnology for Biofuels* **6**: 78.
- Beckwith, T.D., and Light, S.F. (1927) The spirals within the termite gut for class use. *Science* **66**: 656-657.
- Benemann, J.R. (1973) Nitrogen fixation in termites. *Science* **181**: 164-165.
- Berry, D., Mahfoudh, K.B., Wagner, M., and Loy, A. (2011) Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and Environmental Microbiology* **77**: 7846-7849.
- Boucias, D.G., Cai, Y., Sun, Y., Lietze, V.U., Sen, R., Raychoudhury, R., and Scharf, M.E. (2013) The hindgut lumen prokaryotic microbiota of the termite *Reticulitermes flavipes* and its responses to dietary lignocellulose composition. *Molecular Ecology* **22**: 1836-1853.
- Breznak, J.A., and Brune, A. (1994) Role of microorganisms in the digestion of lignocellulose by termites. *Annual Review of Entomology* **39**: 453-487.
- Breznak, J.A., Brill, W.J., Mertins, J.W., and Coppel, H.C. (1973) Nitrogen fixation in termites. *Nature* **244**: 577-580.

- Brulc, J.M., Antonopoulos, D.A., Miller, M.E.B., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A. et al. (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences* **106**: 1948-1953.
- Brune, A. (2014) Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology* **12**: 168-180.
- Brune, A., and Friedrich, M. (2000) Microecology of the termite gut: structure and function on a microscale. *Current Opinion in Microbiology* **3**: 263-269.
- Do, T.H., Nguyen, T.T., Nguyen, T.N., Le, Q.G., Nguyen, C., Kimura, K., and Truong, N.H. (2014) Mining biomass-degrading genes through Illumina-based de novo sequencing and metagenomic analysis of free-living bacteria in the gut of the lower termite *Coptotermes gestroi* harvested in Vietnam. *Journal of Bioscience and Bioengineering* **118**: 665-671.
- Engel, P., Martinson, V.G., and Moran, N.A. (2012) Functional diversity within the simple gut microbiota of the honey bee. *Proceedings of the National Academy of Sciences* **109**: 11002-11007.
- Fröhlich, J., Sass, H., Babenzien, H.-D., Kuhnigk, T., Varma, A., Saxena, S. et al. (1999) Isolation of *Desulfovibrio intestinalis* sp. nov. from the hindgut of the lower termite *Mastotermes darwiniensis*. *Canadian Journal of Microbiology* **45**: 145-152.
- Geib, S.M., Filley, T.R., Hatcher, P.G., Hoover, K., Carlson, J.E., del Mar Jimenez-Gasco, M. et al. (2008) Lignin degradation in wood-feeding insects. *Proceedings of the National Academy of Sciences* **105**: 12932-12937.
- Gilbert, J.A., and Dupont, C.L. (2011) Microbial metagenomics: beyond the genome. *Annual Review of Marine Science* **3**: 347-371.
- Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G. et al. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research* **21**: 494-504.
- He, S., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R.H. et al. (2013) Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS One* **8**: e61126.
- Hongoh, Y. (2011) Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cellular and Molecular Life Sciences* **68**: 1311-1325.
- Hongoh, Y., Sato, T., Dolan, M.F., Noda, S., Ui, S., Kudo, T., and Ohkuma, M. (2007) The motility symbiont of the termite gut flagellate *Caduceia versatilis* is a member of the "Synergistes" group. *Applied and Environmental Microbiology* **73**: 6270-6276.
- Hongoh, Y., Sharma, V.K., Prakash, T., Noda, S., Toh, H., Taylor, T.D. et al. (2008) Genome of an endosymbiont coupling N₂ fixation to cellulolysis within protist cells in termite gut. *Science* **322**: 1108-1109.
- Hopkins, D.W., Chudek, J.A., Bignell, D.E., Frouz, J., Webster, E., and Lawson, T. (1998) Application of ¹³C NMR to investigate the transformations and biodegradation of organic materials by wood- and soil-feeding termites, and a coprophagous litter-dwelling dipteran larva. *Biodegradation* **9**: 423-431.

- Hugenholtz, P., and Tyson, G.W. (2008) Microbiology: Metagenomics. *Nature* **455**: 481-483.
- Hyodo, F., Azuma, J.-I., and Abe, T. (1999) Estimation of effect of passage through the gut of a lower termite, *Coptotermes formosanus* Shiraki, on lignin by solid-state CP/MAS ¹³C NMR. *Holzforschung* **53**: 244-246.
- Imelfort, M., Parks, D., Woodcroft, B.J., Dennis, P., Hugenholtz, P., and Tyson, G.W. (2014) GroopM: an automated tool for the recovery of population genomes from related metagenomes. *PeerJ* **2**: e603.
- Indiragandhi, P., Anandham, R., Madhaiyan, M., Poonguzhali, S., Kim, G.H., Saravanan, V.S., and Sa, T. (2007) Cultivable bacteria associated with larval gut of prothiofos-resistant, prothiofos-susceptible and field-caught populations of diamondback moth, *Plutella xylostella* and their potential for, antagonism towards entomopathogenic fungi and host insect nutrition. *Journal of Applied Microbiology* **103**: 2664-2675.
- Katsumata, K.S., Jin, Z., Hori, K., and Iiyama, K. (2007) Structural changes in lignin of tropical woods during digestion by termite, *Cryptotermes brevis*. *Journal of Wood Science* **53**: 419-426.
- König, H., Li, L., Wenzel, M., and Fröhlich, J. (2006) Bacterial ectosymbionts which confer motility: *Mixotricha paradoxa* from the intestine of the Australian termite *Mastotermes darwiniensis*. In *Molecular Basis of Symbiosis*. Overmann, J. (ed): Springer Berlin Heidelberg, pp. 77-96.
- Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., and Hugenholtz, P. (2008) A bioinformatician's guide to metagenomics. *Microbiology and Molecular Biology Reviews* **72**: 557-578, Table of Contents.
- Leber, T.M., and Balkwill, F.R. (1997) Zymography: A single-step staining method for quantitation of proteolytic activity on substrate gels. *Analytical Biochemistry* **249**: 24-28.
- Li, H., Lu, J., and Mo, J. (2011) Physiochemical lignocellulose modification in the formosan subterranean termite *Coptotermes Formosanus* Shiraki (Isoptera: Rhinotermitidae) and potential uses in the production of biofuels. *BioResources* **7**: 0675-0685.
- Ljungdhal, L. (1986) The autotrophic pathway of acetate synthesis in acetogenic bacteria. *Annual Reviews in Microbiology* **40**: 415-450.
- Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**: D490-D495.
- Mattéotti, C., Haubruge, E., Thonart, P., Francis, F., De Pauw, E., Portetelle, D., and Vandenberghe, M. (2011) Characterization of a new β -glucosidase/ β -xylosidase from the gut microbiota of the termite (*Reticulitermes santonensis*). *FEMS Microbiology Letters* **314**: 147-157.
- Mattéotti, C., Bauwens, J., Brasseur, C., Tarayre, C., Thonart, P., Destain, J. et al. (2012) Identification and characterization of a new xylanase from Gram-positive bacteria isolated from termite gut (*Reticulitermes santonensis*). *Protein Expression and Purification* **83**: 117-127.
- Noda, S., Ohkuma, M., Usami, R., Horikoshi, K., and Kudo, T. (1999) Culture-independent characterization of a gene responsible for nitrogen fixation in the symbiotic microbial community in the gut of the termite *Neotermes koshunensis*. *Applied and Environmental Microbiology* **65**: 4935-4942.

- Odelson, D.A., and Breznak, J.A. (1983) Volatile fatty acid production by the hindgut microbiota of xylophagous termites. *Applied and Environmental Microbiology* **45**: 1602-1613.
- Ohkuma, M., Noda, S., and Kudo, T. (1999) Phylogenetic Diversity of Nitrogen Fixation Genes in the Symbiotic Microbial Community in the Gut of Diverse Termites. *Applied and Environmental Microbiology* **65**: 4926-4934.
- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25**: 1043–1055.
- Peterson, B.F., Stewart, H.L., and Scharf, M.E. (2015) Quantification of symbiotic contributions to lower termite lignocellulose digestion using antimicrobial treatments. *Insect Biochemistry and Molecular Biology* **59**: 80-88.
- Poretsky, R., Rodriguez-R, L.M., Luo, C., Tsementzi, D., and Konstantinidis, K.T. (2014) Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One* **9**: e93827.
- Poulsen, M., Hu, H., Li, C., Chen, Z., Xu, L., Otani, S. et al. (2014) Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proceedings of the National Academy of Sciences* **111**: 14500-14505.
- Raychoudhury, R., Sen, R., Cai, Y., Sun, Y., Lietze, V.U., Boucias, D., and Scharf, M. (2013) Comparative metatranscriptomic signatures of wood and paper feeding in the gut of the termite *Reticulitermes flavipes* (Isoptera: Rhinotermitidae). *Insect Molecular Biology*.
- Scharf, M.E. (2015) Termites as targets and models for biotechnology. *Annual review of entomology* **60**: 77-102.
- Scharf, M.E., and Tartar, A. (2008) Termite digestomes as sources for novel lignocellulases. *Biofuels, Bioproducts and Biorefining* **2**: 540-552.
- Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30.14**: 2068-2069.
- Sethi, A., Slack, J., Kovaleva, E.S., Buchman, G.W., and Scharf, M.E. (2012) Lignin-associated metagene expression in a lignocellulose-digesting termite. *Insect Biochemistry and Molecular Biology*.
- Shen, Z., and Jacobs-Lorena, M. (1997) Characterization of a novel gut-specific chitinase gene from the human malaria vector *Anopheles gambiae*. *Journal of Biological Chemistry* **272**: 28895-28900.
- Strassert, J.F., Desai, M.S., Radek, R., and Brune, A. (2010) Identification and localization of the multiple bacterial symbionts of the termite gut flagellate *Joenia annectens*. *Microbiology* **156**: 2068-2079.
- Suen, G., Weimer, P.J., Stevenson, D.M., Aylward, F.O., Boyum, J., Deneke, J. et al. (2011) The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. *PloS One* **6**: e18814.
- Tartar, A., Wheeler, M., Zhou, X., Coy, M., Boucias, D., and Scharf, M. (2009) Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnology for Biofuels* **2**: 25.

- Tellam, R. (1996) The peritrophic matrix. In *Biology of the insect midgut*: Springer, pp. 86-114.
- Thomas, T., Gilbert, J., and Meyer, F. (2012) Metagenomics-a guide from sampling to data analysis. *Microbial Informatics and Experimentation* **2**: 1-12.
- Todaka, N., Moriya, S., Saita, K., Hondo, T., Kiuchi, I., Takasu, H. et al. (2007) Environmental cDNA analysis of the genes involved in lignocellulose digestion in the symbiotic protist community of *Reticulitermes speratus*. *FEMS Microbiology Ecology* **59**: 592-599.
- Tokuda, G., Tsuboi, Y., Kihara, K., Saitou, S., Moriya, S., Lo, N., and Kikuchi, J. (2014) Metabolomic profiling of ¹³C-labelled cellulose digestion in a lower termite: insights into gut symbiont function. *Proceedings of the Royal Society of London B: Biological Sciences* **281**: 20140990.
- Tokuda, G., and Watanabe, H. (2007) Hidden cellulases in termites: revision of an old hypothesis. *Biology Letters* **3**: 336-339.
- Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T. et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**: 560-565.
- Willner, D., and Hugenholtz, P. (2013) Metagenomics and community profiling: Culture-independent techniques in the clinical laboratory. *Clinical Microbiology Newsletter* **35**: 1-9.
- Yuki, M., Kuwahara, H., Shintani, M., Izawa, K., Sato, T., Starns, D. et al. (2015) Dominant ectosymbiotic bacteria of cellulolytic protists in the termite gut also have the potential to digest lignocellulose. *Environmental Microbiology*.
- Zabel, R.A., and Morrell, J.J. (2012) *Wood microbiology: decay and its prevention*: Academic Press.
- Zhang, X., Matson, E.G., and Leadbetter, J.R. (2011) Genes for selenium dependent and independent formate dehydrogenase in the gut microbial communities of three lower, wood-feeding termites and a wood-feeding roach. *Environmental Microbiology* **13**: 307-323.
- Zhou, Q., Su, X., and Ning, K. (2014) Assessment of quality control approaches for metagenomic data analysis. *Scientific Reports* **4**: 6957.
- Zhou, X., Smith, J.A., Oi, F.M., Koehler, P.G., Bennett, G.W., and Scharf, M.E. (2007) Correlation of cellulase gene expression and cellulolytic activity throughout the gut of the termite *Reticulitermes flavipes*. *Gene* **395**: 29-39.

Chapter 5 A phylogenomic analysis of the bacterial phylum Fibrobacteres

Nurdyana Abdul Rahman¹, Donovan H. Parks¹, Inka Vanwonderghem^{1,2}, Mark Morrison³, Gene W. Tyson¹, Philip Hugenholtz^{1,4,*}

¹Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Queensland, Australia, ²Advanced Water Management Center, Level 4, Gehrman Laboratories Building (60), The University of Queensland, St Lucia, Brisbane, QLD, 4072, Australia, ³The University of Queensland Diamantina Institute, Translational Research Institute, Woolloongabba, QLD, 4102, Australia, ⁴Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland, Australia

Abstract

The Fibrobacteres has been recognised as a bacterial phylum for over a decade, but little is known about the group beyond its environmental distribution, and characterisation of its sole cultured representative genus, *Fibrobacter*, after which the phylum was named. Based on these incomplete data, it is thought that cellulose hydrolysis, anaerobic metabolism, and lack of motility are unifying features of the phylum. There are also contradicting views as to whether an uncultured sister lineage, candidate phylum TG3, should be included in the Fibrobacteres. Recently, chitin-degrading cultured representatives of TG3 were isolated from a hypersaline soda lake and the genome of one species, *Chitinivibrio alkaliphilus*, sequenced and described in detail. Here, we performed a comparative analysis of *Fibrobacter succinogenes*, *C. alkaliphilus* and eight near or substantially complete Fibrobacteres/TG3 genomes of environmental populations recovered from termite gut, anaerobic digester, and sheep rumen metagenomes. We propose that TG3 should be amalgamated with the Fibrobacteres phylum based on robust monophyly of the two lineages and shared character traits. Polymer hydrolysis, using a distinctive set of glycoside hydrolases and binding domains, appears to be a prominent feature of members of the Fibrobacteres. Not all members of this phylum are strictly anaerobic as some termite gut Fibrobacteres have respiratory chains adapted to the microaerophilic conditions found in this habitat. Contrary to expectations, flagella-based motility is predicted to be an ancestral and common trait in this phylum and has only recently been lost in *F. succinogenes* and its relatives based on phylogenetic distribution of flagellar genes. Our findings extend current understanding of the Fibrobacteres and provide an improved basis for further investigation of this phylum.

5.1 Introduction

The phylum Fibrobacteres is recognised as a major line of descent in the bacterial domain but is understudied due to limited representation by axenic cultures. The only described genus in this lineage is *Fibrobacter* (Montgomery et al., 1988, originally classified as *Bacteroides* Hungate, 1950), after which the phylum was named (Ludwig and Klenk, 2001). *Fibrobacter* currently comprises two species, *F. succinogenes* isolated from a cow rumen (Hungate, 1950) and *F. intestinalis* isolated from a rat cecum (Montgomery and Macy, 1982), of which the former has a publicly available genome sequence (Suen et al., 2011). Both species are primary degraders of cellulosic plant biomass in herbivore guts (Hungate, 1950; Montgomery et al., 1988), which has prompted the suggestion that cellulose degradation may be a unifying feature of the phylum (Ransom-Jones et al., 2012; Ransom-Jones et al., 2014). This is supported by culture-independent 16S rRNA-based environmental surveys identifying relatively high numbers of diverse members of the Fibrobacteres in cellulolytic ecosystems (Ransom-Jones et al., 2012; Ransom-Jones et al., 2014).

Candidate phylum TG3 (Termite group 3) is often phylogenetically associated with the Fibrobacteres based on comparative analyses of the 16S rRNA gene (Hongoh et al., 2005; Hongoh et al., 2006; Warnecke et al., 2007; He et al., 2013; Sorokin et al., 2014). TG3 was initially detected in environmental surveys of termite guts (Hongoh et al., 2005), but was later found to be present in a diverse range of habitats (Hongoh et al., 2006). Recently, the first isolates for TG3 have been described (Sorokin et al., 2012), one of which has been named *Chitinivibrio alkaliphilus* and its genome sequenced (Sorokin et al., 2014). *C. alkaliphilus* is a haloalkaliphilic anaerobic chitin-utilising bacterium isolated from soda lake sediments. There have been conflicting views as to whether TG3 should be merged with the Fibrobacteres or retained as a separate phylum (Sorokin et al., 2014).

Recent developments in metagenomics provide the opportunity to obtain genomic representation of uncultured Fibrobacteres and TG3 populations which can be used to evaluate conservation of polymer (cellulose and chitin) degradation and other metabolic properties across these lineages, and the robustness of the association between the two phyla. Here, we used differential coverage binning (Albertsen et al., 2013) to obtain seven Fibrobacteres and one TG3 population genomes from termite gut, sheep rumen and anaerobic digester samples. This substantially expands the genomic coverage of both groups and comparative analyses of these genomes with the publicly available *F. succinogenes* and *C. alkaliphilus* genomes suggest that polymer hydrolysis is a phylogenetically widespread trait in these lineages. We propose that candidate phylum TG3 should

be classified as part of the Fibrobacteres based on shared character traits and phylogenetic analyses of concatenated gene sets supporting a robust association between the two groups.

5.2 Methods

5.2.1 Samples and metagenome sequencing

DNA samples described in previous 16S rRNA community profiling studies were used in the present study for shotgun sequencing. These comprised four termite samples; MC05, MC06, MC07 and IN01 (Rahman et al., 2015) and six anaerobic digester samples taken from 3 reactors (AD1-3) at two time points (day 96 and 362) (Vanwonderghem et al., 2014). A publicly available sheep rumen metagenome (BioProject acc. PRJNA214227) was also included in the study together with two reference genomes; *Fibrobacteres succinogenes* S85 (BioProject acc. PRJNA41169) and *Chitinivibrio alkaliphilus* ACht1 (BioProject acc. PRJNA195589). Shotgun libraries were prepared using the Nextera XT Sample Preparation Kit (or TruSeq DNA Sample Preparation Kits v2 for AD1-3 day 96) (Illumina, San Diego, CA, USA) and library DNA concentrations were measured using the QuantIT kit (Molecular probes, Carsbad, CA, USA) and equimolar-pooled for sequencing. Between a quarter and a third of an Illumina HiSeq 2000 flowcell of paired-end sequences (2 x 100bp with an average fragment size of 320) were obtained for each library.

5.2.2 Sequence assembly and population genome binning

For the termite datasets, paired-end reads were merged and adaptors removed using SeqPrep v1.1 (<https://github.com/jstjohn/SeqPrep>), and then quality trimmed with a Q-value of 20 using Nesoni v0.128 (<http://www.vicbioinformatics.com/software.nesoni.shtml>). Adaptor removal and quality trimming was performed using CLC Workbench v6 (CLC Bio, Taipei, Taiwan) for the anaerobic digester (AD) datasets (Vanwonderghem et al., *in prep*). *De novo* assemblies of the termite and AD datasets were generated using CLC Workbench v6 using a word size of 63 and a minimum contig length of ≥ 500 bp. Reads from each samples were mapped to the assembled contigs using the BWA-MEM algorithm in BWA v0.5.5 with default parameters (Li, 2013). Population genomes were obtained using the differential coverage binning method of GroopM (Imelfort et al., 2014) with default parameters. The termite and AD metagenomes were binned independently using GroopM v0.1 and v0.2, respectively. Briefly, reads from each sample were mapped onto their corresponding co-assemblies and coverage patterns for each scaffold were calculated, transformed, and projected onto a 3-dimensional space in which scaffolds from the same population genome cluster together (Imelfort et al., 2014). Manual refinement of selected genomes was performed using the GroopM refine function in order to merge bins with compatible genome characteristics

(i.e., GC and coverage statistics) and split bins that appeared to be aggregates of two or more genomes. For the sheep rumen metagenome, population genomes were recovered using a distribution-based binner (DBB v1.0.1; <https://github.com/dparks1134/DBB>) since multiple related samples were not available for differential coverage binning. This method identified contigs likely to belong to the same population based on the GC-content, tetranucleotide signature, and coverage of individual contigs. Genome completeness and contamination was estimated using the lineage-specific marker sets determined by CheckM v1.0.3 (Parks et al., 2015).

5.2.3 Taxonomic assignment of population genomes

Population genomes estimated to be >60% complete and <10% contaminated were placed in a maximum likelihood tree of 2,358 reference genomes based on a concatenation of 83 marker genes as described previously (Soo et al., 2014). The inferred phylogeny was used to identify putative members of the Fibrobacteres and TG3 lineages. To corroborate genome-based identifications, 16S rRNA genes or gene fragments associated within the population genomes were identified with CheckM (Parks et al., 2015) and aligned with reference Fibrobacteres and TG3 sequences obtained from SILVA database release 119 (Quast et al., 2012) using ssu-align v0.1 (Nawrocki et al., 2009). Poorly represented leading and trailing columns of the multiple sequence alignment were manually trimmed, and a maximum likelihood tree inferred with FastTree v2.1.7 (Price et al., 2009). Sequences greater than 1200 nt were selected for the purposes of calculating nonparametric bootstrap support values. These selected sequences were reanalysed using FastTree followed by 100 bootstrap replicates, and support values propagated to the full tree consisting of both short and long sequences. Phylogenetic tree and bootstraps values were scaled and edited in ARB (Ludwig et al., 2004) and Adobe Illustrator CS6 (Adobe). All Fibrobacteres/TG3 population genomes have been deposited at JGI IMG/ER under the accessions 2522572000, 2522572002, 2522572004, 2522572005, 2582580742, 2582580743, 2585427501, 2606217802 and GenBank/DDBJ/EMBL as individual Biosamples under the multispecies BioProject PRJNA293241.

5.2.4 Genome annotation and metabolic reconstruction

The draft Fibrobacteres and TG3 genomes were uploaded to the Integrated Microbial Genomes with Microbiome Samples-Expert Review (IMG/ER) system (Markowitz et al., 2014) for automated annotation with IMG/M Metagenome Gene Calling. KEGG pathway maps were visualised by uploading KEGG (Kyoto Encyclopedia of Genes and Genomes) annotations to the KEGG Mapper - Colour Pathway (http://www.genome.jp/kegg/tool/map_pathway3.html). Glycoside hydrolases (GHs) and carbohydrate-binding modules (CBMs) were identified using the CAZy database (Lombard et al., 2014) via dbCAN (Yin et al., 2012). Signal peptide predictions were performed

using SignalP (Petersen et al., 2011). IMG/ER identified methyl-accepting proteins were scanned for chemotaxis protein domain using InterProScan5 (Jones et al., 2014). The draft genomes were also annotated with PROKKA v1.7 using default settings (Seemann, 2014). The final gene and pathway inventories of the putative Fibrobacteres and TG3 genomes were based on a combination of the IMG and PROKKA annotations and functional classifications based on COG (Clusters of Orthologous Groups), KO, Enzyme, Pfam, or TIGRFam assignments. Metabolic reconstructions based on these inventories were prepared in Adobe Illustrator CS6 (Adobe).

5.2.5 Genome and protein family comparative analyses

Average amino acid identities (AAI) between homologues in genome pairs were calculated using the AAI calculator with default settings in CompareM v0.0.4 (<https://github.com/dparks1134/CompareM>). Heatmaps of the relative abundance of genes and pathways within genomes were generated with STAMP v2.0.9 (Parks et al., 2014). Phylogenetic analysis of selected proteins (GHs, CBMs, cytochrome bd, fibro-slime domain, flagellar proteins) in the population genomes was performed by identifying homologues within IMG v4.510 (Markowitz et al., 2014) using BLASTP. A gene was considered homologous if it had an expectation value $\leq 1e-5$, an amino acid identity $\geq 50\%$, and an alignment length of $\geq 30\%$. Proteins alignments were obtained using MAFFT v7.221 (Standley, 2013) and trees inferred using FastTree v2.1.7 under the WAG+G models and support values determined using 100 non-parametric bootstrap replicates.

5.3 Results and Discussion

5.3.1 Recovery of population genomes from environmental metagenomic datasets

Bulk DNAs extracted from termite whole gut samples for 16S rRNA-based community profiling (Rahman et al., 2015) were used in the present study. A total of 74 Gb of Illumina 2 x 100 bp data were sequenced from four sets of *Microcerotermes* whole gut samples (30 guts per set) obtained from the same nest, IN01, in Brisbane, Queensland. Similarly, 71 Gb was sequenced from three sets of *Nasutitermes* whole gut samples (30 guts per set) collected from three mounds within a 1 km radius in Murphy's Creek, South East Queensland (MC05, MC06, MC07). Bulk DNAs extracted from three lab-scale anaerobic digesters collected at two timepoints (AD1 to 3; reported in Vanwonterghem et al. (2014)) were also sequenced to produce a total of 111 Gb (2 x 100 bp Illumina reads). Publicly deposited metagenomic datasets were also screened for the presence of Fibrobacteres genomes (data not shown), of which one, a sheep rumen microbiome (BioProject acc. PRJNA214227, SRR948090; 9.9 Gb of 2 x 100 bp Illumina reads) produced a genome of sufficient quality for comparative analysis. Sequence datasets from each habitat were independently

assembled and binned (Appendix D: **Table S5.1**). A total of 303 population genomes with >60% completeness and <10% contamination (estimated by CheckM; Parks et al. (2015)) were obtained from the four sample types and, of these, eight were phylogenetically affiliated with the publicly available isolate genomes of Fibrobacteres (*Fibrobacter succinogenes* S85, acc. PRJNA41169) and TG3 (*Chitinivibrio alkaliphilus* AChT1, acc. PRJNA195589) (**Figure 5.1**). All eight genomes had low contamination, four were near complete and four were substantially complete according to CheckM estimates and classification (Parks et al., 2015). Together with the two reference organisms, genome size and GC content range from 2.4 to 3.8 Mb and 37.4 to 53.9% respectively (**Table 5.1**) comparable to other phyla of similar phylogenetic breadth (Lightfield et al., 2011).

5.3.2 An expanded phylogenetic classification of the phylum Fibrobacteres

We constructed a phylogenetic tree based on a concatenated alignment of 83 bacterial single copy marker genes (Dupont et al., 2012). The ingroup comprised the two complete reference genomes representing the Fibrobacteres (Suen et al., 2011) and TG3 (Sorokin et al., 2014) lineages and eight population genomes obtained in this study (**Table 5.1**). We evaluated the monophyly of these genomes using an outgroup consisting of 2358 genomes from 33 phyla. The Fibrobacteres and TG3 genomes formed a robustly monophyletic group (**Figure 5.1A**) supporting the previously noted relationship between these lineages (Hongoh et al., 2006; Warnecke et al., 2007; Krieg et al., 2011; He et al., 2013; Mikaelyan et al., 2015). Therefore, we propose to amalgamate TG3 as one or more classes within the phylum Fibrobacteres based on this robust phylogenetic association and shared character traits described below. Additionally, all 10 ingroup genomes contain signature inserts in their RNA polymerase β' subunit and serine hydroxymethyltransferase genes that identify them as members of the FCB superphylum (Gupta, 2004). Using the partial 16S rRNA gene sequences identified in a number of the population genomes (**Table 5.1**), we placed the genomes in the broader context of the 16S rRNA-defined Fibrobacteres lineage (**Figure 5.1B**). *F. succinogenes* S85, AD_80, AD_111 and SR_36 belong to the family *Fibrobacteraceae* (Spain et al., 2010), with AD_312 likely representing a separate family in the same order (*Fibrobacterales*) based on AAI similarities (Appendix D: **Table S5.2**; Konstantinidis and Tiedje (2005)). IN01_31, IN01_221 and IN01_307 form a monophyletic cluster found exclusively in termite guts previously referred to as candidate order TFG-1 (Warnecke et al., 2007). We propose the candidatus name, *Fibromonas termitidis*, for the most complete of these genomes, IN01_221, and the family and order names, *Fibromonadaceae* and *Fibromonadales* for this group and related 16S rRNA sequences (**Figure 5.1**). Unfortunately, population genome MC_77 lacked a 16S rRNA sequence so could not be placed within the 16S framework. However, it likely belongs to the TG3 lineage, and more specifically in the termite cluster proximate to isolate AChT6-1 (**Figure 5.1**; Sorokin et al. (2012)).

Sorokin et al., (2014) proposed the class *Chitinivibrionia* to accommodate *C. alkaliphilus* ACht1, which now becomes the second recognised class within the Fibrobacteres due to its amalgamation with TG3 (**Figure 5.1**). We have provisionally included MC_77 in the class *Chitinivibrionia*, however, given the depth of the relationship with *C. alkaliphilus* (**Figure 5.1**), MC_77 and isolate ACht6-1 may represent a distinct class within the expanded phylogenetic representation of the Fibrobacteres phylum.

Table 5.1: Summary statistics of Fibrobacteres genomes.

Genomes	Source	Estimated population genome size (Mb)	GC (%)	No. of contigs	Estimated Completeness ¹ (%)	Estimated Contamination ¹ (%)	No. of genes	rRNA genes	Coding Density (%)	Refs
PURE CULTURE										
<i>Fibrobacter succinogenes</i> S85	Cow rumen	3.8	48.1	1	100.0	2.5	3188	5S,16S,23S	91.3	Suen <i>et al.</i> , 2008
<i>Chitinivibrio alkaliphilus</i> ACh1	Soda Lake	2.6	46.2	99	97.9	0.0	2346	5S,16S,23S	93.1	Sorokin <i>et al.</i> , 2014
POPULATION GENOMES										
AD_80	Bioreactor	3.3	51.4	123	88.9	1.7	2801	5S	91.1	Present study
AD_111	Bioreactor	3.6	50.2	189	88.1	0.2	3069	5S,16S,23S	91.3	Present study
SR_36	Sheep rumen	3.4	53.9	50	99.2	1.7	2906	5S	93.1	PRJNA214227 ²
AD_312	Bioreactor	2.8	37.4	56	100.0	3.5	2392	-	90.9	Present study
IN01_31	<i>Microcerotermes</i> gut	3.2	43.2	220	96.6	2.6	3417	5S,23S	88.5	Present study
<i>Fibromonas termitidis</i> ³	<i>Microcerotermes</i> gut	3.2	43.1	211	98.3	1.7	3355	16S	90.7	Present study
IN01_307	<i>Microcerotermes</i> gut	2.6	41.5	157	87.9	1.7	2779	-	92.4	Present study
MC_77	<i>Nasutitermes</i> gut	2.4	52.4	317	73.3	0.0	2324	-	85.2	Present study

¹Estimated completeness and contamination based on lineage-specific single copy marker genes (Parks *et al.*, 2015)

²Bioproject accession number

³IN01_221

Table 5.2: Inventory of glycoside hydrolases (GHs) identified in the Fibrobacteres genomes, organised by functional category.

CAZy family	Known activity	pfam domain	Fibrobacteria								Chitinivibrionia		Cellulolytic bacteria (35 genomes ¹)	Bacterial average across 2038 genomes		
			Fibrobacteraceae				Fibromonadaceae				<i>C. alkali-philus</i> ACh1	MC_77		Avg	Avg	Prevalence (%)
			<i>F. succinogenes</i> S85	AD_80	AD_111	SR_36	AD_312	IN01_31	IN01_221	IN01_307						
Cellulases																
GH5	cellulase	PF00150	13.0	16.7	16.7	15.2	18.8	27.5	18.0	15.2	6.3	23.3	2.37	1.46	30.9	17
GH9	endoglucanase	PF00759	9.0	10.3	10.4	8.7	5.8	10.0	12.8	12.1	18.8	6.7	0.91	0.31	9.6	15
GH44	endoglucanase	PF12891	1.0	1.3	1.0	1.1	1.5	-	-	-	-	-	0.03	0.03	2.3	7
GH45	endoglucanase	PF02015	4.0	3.9	5.2	4.4	2.9	2.5	7.7	9.1	-	3.3	0.02	0.01	0.3	3
Subtotals (%)			27.0	32.1	33.3	29.4	29.0	40.0	38.5	36.4	25.0	33.3	3.33	1.81		
Chitinases																
GH18	chitinase	PF00704	2.0	2.6	2.1	2.2	2.9	2.5	2.6	3.0	3.1	3.3	3.12	2.16	30.0	19
GH19	chitinase	PF00182	-	-	-	-	-	-	-	-	6.3	-	0.55	0.32	9.7	10
GH20	β -hexosaminidase	PF00728	-	-	-	-	-	-	-	-	-	3.3	1.15	0.91	23.2	17
Subtotals (%)			2.0	2.6	2.1	2.2	2.9	2.5	2.6	3.0	9.4	6.7	4.82	3.39		
Hemicellulases																
GH8	endo-xylanases	PF01270	6.0	3.9	7.3	7.6	5.8	5.0	5.1	3.0	15.6	3.3	0.81	0.56	20.1	17
GH10	endo-1,4- β -xylanase	PF00331	8.0	5.1	7.3	6.5	2.9	2.5	2.6	3.0	6.3	-	1.06	0.43	12.7	14
GH11	xylanase	PF00457	4.0	1.3	4.2	5.4	1.5	2.5	5.1	3.0	-	-	0.31	0.08	4.3	11
GH26	β -mannanase & xylanase	PF02156	4.0	3.9	4.2	3.3	2.9	-	-	3.0	-	-	0.98	0.16	6.9	11
GH53	endo-1,4- β -xylanase	PF07745	2.0	1.3	1.0	2.2	1.5	2.5	2.6	3.0	-	-	0.55	0.22	10.6	14
Subtotals (%)			24.0	15.4	21.0	25.0	14.5	12.5	15.4	15.2	21.9	3.3	3.72	1.45		
Debranching enzymes																
GH16	xyloglucanases & xyloglycosyltransferases	PF00722	4.0	5.1	4.2	3.3	4.4	0.0	0.0	0.0	3.1	0.0	1.72	0.78	19.7	20
GH74	endoglucanases & xyloglucanases	-	2.0	3.9	2.1	2.2	2.9	0.0	0.0	0.0	0.0	0.0	0.27	0.10	4.3	9
GH51	alpha-L-arabinofuranosidase	-	1.0	1.3	1.0	1.1	1.5	0.0	2.6	3.0	0.0	0.0	0.78	0.41	15.0	14
GH54	alpha-L-arabinofuranosidase	PF09206	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.02	0.9	6
Subtotals (%)			8.0	10.3	8.3	6.5	8.7	0.0	2.6	3.0	3.1	0.0	2.77	1.31		
Oligosaccharide-degrading enzymes																
GH1	β -glucosidase & other β -linked dimers	PF00232	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.1	3.3	2.73	3.66	46.6	17
GH2	β -galactosidases and other β -linked dimers	PF00703 PF02836 PF02837	2.0	2.6	2.1	2.2	2.9	0.0	0.0	3.0	3.1	0.0	2.88	1.92	38.2	18
GH3	mainly β -glucosidases	PF00933 PF01915	3.0	2.6	2.1	1.1	2.9	5.0	0.0	3.0	0.0	0.0	4.68	5.87	73.4	27
GH43	arabinases & xylosidases	PF04616	14.0	11.5	9.4	13.0	13.0	5.0	5.1	0.0	0.0	3.3	2.82	1.30	22.2	15

Table 5.2: Continued.

CAZy family	Known activity	pfam domain	Fibrobacteria								Chitinivibronia		Cellulolytic bacteria (35 genomes ¹)	Bacterial average across 2038 genomes		
			Fibrobacteraceae				Fibromonadaceae				<i>C. alkali-philus</i> AChT1	MC ₇₇		Avg	Avg	Prevalence (%)
			<i>F. succinogenes</i> S85	AD ₈₀	AD ₁₁₁	SR ₃₆	AD ₃₁₂	IN01 ₃₁	IN01 ₂₂₁	IN01 ₃₀₇			Avg			
GH52	β-xylosidase	PF03512	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.3	0.04	0.03	1.5	5
Subtotals (%)			19.0	16.7	13.5	16.3	18.8	10.0	5.1	6.1	6.2	10.0	13.14	12.79		
Cellobiose/chitobiose phosphorylase																
GH84	cellobiose/chitobiose phosphorylase	PF06165 PF06205	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.1	0.0	0.23	0.16	4.8	7
GH94	cellobiose/chitobiose phosphorylase	PF06165 PF06205	1.0	1.3	1.0	1.1	1.5	5.0	2.6	3.0	6.3	3.3	0.58	0.66	18.8	18
Subtotals (%)			1.0	1.3	1.0	1.1	1.5	5.0	2.6	3.0	9.4	3.3	0.81	0.82		
Others																
GH13	unknown lysozyme type	PF02903	3.0	3.9	3.1	3.3	5.8	7.5	10.3	12.1	6.3	6.7	16.52	16.67	80.2	27
GH23	G/peptidoglycan lyase/chitinase	-	3.0	3.9	3.1	3.3	4.4	12.5	10.3	12.1	12.5	20.0	8.01	17.21	80.7	30
GH27	unknown		1.0	1.3	1.0	1.1	1.5	0.0	0.0	0.0	0.0	0.0	0.44	0.09	5.1	10
GH30	glucosylceramidase	PF02055	4.0	3.9	4.2	4.4	2.9	0.0	2.6	0.0	0.0	0.0	0.52	0.31	11.6	15
GH57	alpha-amylase	PF03065	3.0	3.9	2.1	2.2	2.9	5.0	2.6	3.0	3.1	10.0	0.70	2.43	22.8	27
GH77	amylomaltase or 4-α-glucanotransferase	PF02446	1.0	1.3	1.0	1.1	1.5	2.5	2.6	3.0	3.1	3.3	1.47	3.20	55.0	26
GH95	α-1,2-L-fucosidase/α-L-fucosidase	-	1.0	1.3	1.0	1.1	1.5	0.0	0.0	0.0	0.0	0.0	0.35	0.22	9.8	14
GH105	unknown	PF07470	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.77	0.31	12.2	14
GH106	α-L-rhamnosidase	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.3	0.16	0.09	4.5	8
GH109	α-N-acetylgalactosaminidase	-	0.0	0.0	0.0	1.1	1.5	0.0	0.0	0.0	0.0	0.0	5.81	5.61	65.1	27
GH116	unknown		1.0	1.3	1.0	1.1	1.5	2.5	2.6	3.0	0.0	0.0	0.04	0.21	4.7	9
GH127	β-L-arabinofuranosidase	-	1.0	1.3	1.0	1.1	1.5	0.0	2.6	0.0	0.0	0.0	0.77	0.32	14.7	14
Subtotals (%)			19.0	21.8	17.7	19.6	24.7	30.0	33.3	33.3	25.0	43.3	35.51	46.67		
Total estimated glycoside hydrolases (glycoside hydrolases with signal peptide (%))			100	78	96	92	69	40	39	33	32	30	-	-		
			71.8	64.1	72.1	76.4	48.7	35.6	46.5	40.0	31.4	28.6	-	-		
Total estimated genes (signal peptide to total genes (%))			2871	2754	3008	2855	2344	3391	3347	2773	2304	2321	4512 ³	3395 ³		
			21.2	17.9	21.6	22.0	15.2	10.8	11.1	11.5	7.4	7.7				
% of genes that are glycoside hydrolases			3.5	2.8	3.2	3.2	2.9	1.2	1.2	1.2	1.4	1.3	1.2⁴	0.9⁴		

¹Based on Koeck et al. (2014)

²Of a total of 30 bacterial phylum

³Estimated average genes across 35 and 2038 bacterial genomes respectively

⁴Average of 116 GHs to total genes of 35 and 2038 bacterial genomes respectively

5.3.3 Inferred metabolism of Fibrobacteres genomes

We performed comparative analyses of the two isolate and eight draft population genomes (**Table 5.1**) to infer metabolic properties associated with the *Fibrobacteres* in the context of their environmental settings.

5.3.3.1 Polymer hydrolysis

Cellulases

Members of the Fibrobacteres are best known for their ability to hydrolyse plant polymers in anoxic habitats such as the bovine rumen (Suen et al., 2011; Jewell et al., 2013; Ransom-Jones et al., 2014) and termite gut (Warnecke et al., 2007; He et al., 2013). Therefore, we began by identifying genes encoding glycoside hydrolases (GHs) classified according to the CAZy database (Lombard et al., 2014). All ten genomes contained numerous GHs representing between 1.2 to 3.5% of the total genes, which is higher than the bacterial average of 0.9%, but similar to other cellulolytic bacteria (2%) (**Table 5.2**). However, polymer-degrading enzymes are highly over-represented in the Fibrobacteres GH inventory relative to other recognised cellulolytic bacteria (cellulases - 25% vs 3%, xylanases - 15% vs 4%). The proportion of Fibrobacteres GHs with signal peptides is also much higher than that for the average Fibrobacteres gene (28.6 ± 76.4 vs 7.4 ± 21.6 respectively) which is as expected for proteins involved in extracellular deconstruction of carbohydrate polymers (Lombard et al., 2014).

A quarter of the GHs in the *Fibrobacteraceae* and *Chitinivibrionia* and over a third of the GHs in the *Fibromonadaceae* are cellulases. Most of the cellulases belong to families GH5 and GH9 which are widely distributed in bacteria (**Figure 5.2**) (present in $\geq 50\%$ of recognised phyla; **Table 5.2**; Berlemont and Martiny (2013)). The less common cellulase family GH45, previously noted to be distinctive of *F. succinogenes* (Suen et al., 2011; Dai et al., 2012) and related organisms in the termite hindgut (Warnecke et al., 2007), is present in all studied representatives of the Fibrobacteres, with the exception of *C. alkaliphilus* (**Figure 5.2** and **Table 5.2**). Cellulase family GH44 is distinctive of the *Fibrobacteraceae* in the context of the Fibrobacteres although it has been identified in members of six other bacterial phyla. The previously noted absence of the classical exo-acting β -1,4 glucanase families GH6, GH7 and GH48 in *F. succinogenes* is upheld across the phylum supporting the hypothesis that the Fibrobacteres have a distinctive suite of carbohydrate-active enzymes and lignocellulose hydrolysis mechanism (Morrison et al., 2009, Wilson 2009). Furthermore, the distinctive basic terminal domain (~80 AA in the C-terminus) noted in *F. succinogenes* cellulases (Iyo and Forsberg, 1996; Malburg et al., 1996; Qi et al., 2007, 2008) is

widespread in cellulases of all members of the Fibrobacteres. Cellulases play an important role in the habitats from which the Fibrobacteres genomes were obtained (**Table 5.1**) with the possible exception of the soda lake from which *C. alkaliphilus* was recovered. Although *C. alkaliphilus* encodes a high proportion of cellulases relative to the bacterial average (**Table 5.2**), it was reported to be unable to grow on cellulose as a sole carbon source (Sorokin et al., 2014) indicative of their role being relevant to polymer deconstruction rather than energy acquisition.

Hemicellulases and debranching enzymes

As with the cellulases, hemicellulases and debranching enzymes are present in the Fibrobacteres genomes at much higher relative abundance than the bacterial average, (12.5 to 32.3% vs 2.7%) with the exception of MC_77 (3.3%; **Table 5.2**). Five hemicellulase families, primarily endoxylanases, were identified in the eight *Fibrobacteria* genomes, likely reflecting the importance of xylan hydrolysis in animal and insect gut ecosystems (Allgaier et al., 2010; Tokuda et al., 2014). Debranching enzymes, responsible for cleaving the side chains (glycosidic and/or ester linkages) from xylan backbones (Sethi and Scharf, 2013), were most prevalent in the *Fibrobacteraceae*. Families GH51 and GH54 are most commonly alpha-L-arabinofuranosidases responsible for removing arabinose side chains from xylan which is an important constituent of plant lignocellulose (He et al., 2013). GH51s were common in the *Fibrobacteria*, whereas GH54 was only identified in *F. succinogenes* S85 and AD_111 (**Table 5.2**), despite being closely related to the AD_80 and SR_36 population genomes (**Figure 5.1**).

Chitinases

There are three GH families with recognised chitinase activity, GH18, GH19 and GH20, the first two of which are responsible for hydrolysis of insoluble chitin to soluble oligosaccharides in the periplasm (LeClerc et al., 2007; Beier and Bertilsson, 2013). GH20 hydrolyses N-Acetylglucosamine (GlcNAc) molecules from chitin oligomers (Beier and Bertilsson, 2013) or directly from chitin polymers (LeClerc et al., 2007). As expected, *C. alkaliphilus* has the highest proportion of chitinases, approximately three times the bacterial average (**Table 5.2**), as it is a chitin-degrading specialist (Sorokin et al., 2014). Furthermore, it has two types of chitinases, GH18 and GH19, which has been postulated to improve substrate degradation due to synergistic enzyme interactions (Beier and Bertilsson, 2013). The closest phylogenetic neighbour of *C. alkaliphilus* in this study, MC_77, similarly has representatives of two chitinase families (GH18 and GH20) and a higher than average proportion of chitinases (**Table 5.2**) suggesting that chitin degradation may be occurring in the termite hindgut from which the MC_77 genome was obtained. Chitinases have rarely been considered in the context of Fibrobacteres, however, all representatives of this phylum

had GH18-encoding genes at the bacterial average (**Table 5.2**), indicating the potential for this function in primarily lignocellulose-degrading gut communities.

Accessory attachment genes for polymer degradation

The adhesion of cellulolytic anaerobic bacterial to plant biomass is considered a prerequisite step in breaking down plant cell walls (Morrison and Miron, 2000; Miron et al., 2001). In anaerobic gut bacteria such as *F. succinogenes* and *Ruminococcus* species, surface-associated cellulolytic enzymes complexes (cellulosomes, Dassa et al. (2014)), individual GHs possessing non-catalytic carbohydrate-binding modules (CBM, Qi et al. (2005)), and Type IV pilin like structures (Pegden et al., 1998) are known to be responsible for adhesion. Fifteen CBM families are represented in the *Fibrobacteria* and *Chitinivibrionia* genomes mostly targeting cellulose, hemicellulose or chitin (**Table 5.3**) which is consistent with the GH profiles (**Table 5.2**). There are approximately four times as many CBMs in the *Fibrobacteraceae* as in *Fibromonadaceae* and *Chitinivibrionia*, which is also broadly consistent with the relative abundances of GHs in these groups. The CBM families also showed lineage-specific patterns. For example both CBM6 and CBM35 are all overrepresented in the *Fibrobacteraceae* compared to the *Fibromonadaceae*, but the opposite is apparent for CBM11, CBM32 and CBM50 (**Table 5.3**). This suggests that CBMs in the Fibrobacteres have most often been vertically inherited and have not been distributed between lineages by horizontal transfer. This is supported by phylogenetic reconstruction of the Fibrobacteres CBMs which shows mostly vertical transmission and in some lineages expansion of families via gene duplication (Appendix D: **Figure S5.1**). Higher relative abundances of certain CBM families also correlate with the observed differences in GH family abundances. For example, CBM6 is often associated with the hemicellulose-associated families, GH10 and GH43 (Suen et al., 2011), and all three of these families are overrepresented in the *Fibrobacteraceae* relative to the *Fibromonadaceae* (**Table 5.2** and **Table 5.3**).

As previously reported for *F. succinogenes* (Suen et al., 2011), no clostridial-like cohesin or dockerin-like modules were identified in any of the Fibrobacteres genomes, indicative of an absence of cellulosomes in this lineage. Two other putative cellulose binding proteins have been reported in *F. succinogenes*; TIGR02145 and TIGR02148 (Morrison et al., 2009; Suen et al., 2011). TIGR02145 is a domain of ~175 to 200 amino acids with an inferred extracytoplasmic location, and has been suggested to be a possible cohesin analogue (Warnecke et al., 2007). It is present in high copy number in all of the Fibrobacteres genomes (17 to 119 copies) with the exception of *C. alkaliphilus*. TIGR02148 is a fibro-slime domain-containing protein originally identified in the *F. succinogenes* genome and implicated in adherence to plant biomass (Toyoda et al., 2009). We

found this protein family to be present in all Fibrobacteres genomes, again with the exception of *C. alkaliphilus* (**Table 5.4**). Therefore, these putative adhesion proteins are not only distinctive of *F. succinogenes*, but of the Fibrobacteres phylum as a whole. A phylogenetic reconstruction of the fibro-slime protein family indicates multiple duplication events in the class *Fibrobacteria* resulting in up to 10 copies per genome (**Table 5.4** and Appendix D: **Figure S5.2**). Interestingly, one of the two fibro-slime proteins identified in the termite *Chitinivibrionia* genome, MC_77, contains a flagellar domain (*flgD*) suggesting that polymer attachment in this species may be flagella-mediated. Type IV pili are known to facilitate attachment of *F. succinogenes* cells to cellulose (Qi et al., 2007) and Gram negative cells to chitin (Li et al., 2003; Giltner et al., 2012). All Fibrobacteres genomes contain the necessary genes for synthesis of Type IV pili (**Table 5.4**) suggesting that this may be a widespread auxiliary mechanism used by members of this phylum to attach to polymers, and perhaps, to facilitate a “twitching” motility phenotype.

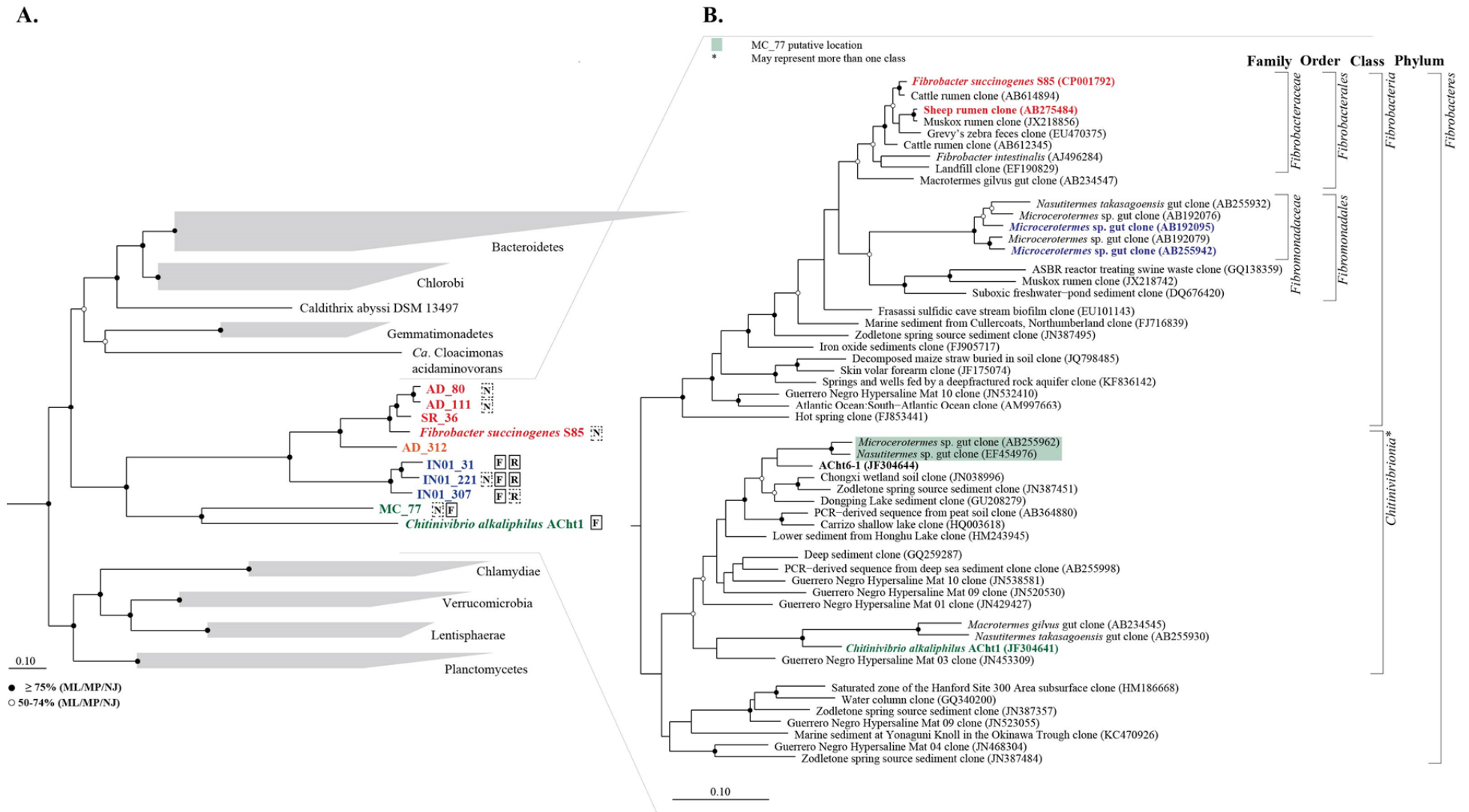


Figure 5.1: Phylogenetic analysis of the phylum Fibrobacteres. (A) Maximum likelihood tree of the phylum Fibrobacteres based on alignment of 83 concatenated proteins as previously described (Soo et al, 2014). The tree was inferred using an outgroup comprising 2358 genomes from 33 phyla. For clarity, only the immediate phylum-level neighbourhood of the Fibrobacteres is shown. Fibrobacteraceae genomes are shown in red; Fibromonadaceae in blue; and Chitinivibronia in green. Fibrobacteres genomes encoding nitrogen-fixing, flagellar and/or respiratory genes are indicated by N, F and R in boxes (dotted box indicates incomplete genes), respectively. Bootstrap support for interior nodes using multiple inference methods is shown according

to the legend at the lower left of the fig; ML=Maximum Likelihood, MP= Maximum Parsimony, NJ=Neighbour Joining. (B) Maximum likelihood tree based on 16S rRNA genes from Fibrobacteres and TG3 obtained from SILVA database release 119 (Quast et al., 2013). The closest matches to the partial 16S rRNA sequences obtained from the population genomes are indicated by colour matching to panel A, noting that the position of MC_77 is estimated since this genome lacks a 16S rRNA sequence. Isolates are bolded in black. Taxonomic group names by rank are proposed to the right of the tree, also see main text. Node support values are as described for panel A.

Table 5.3: Inventory of accessory attachment genes for polymer hydrolysis identified in the Fibrobacteres genomes, organised by Carbohydrate-binding modules (CBMs).

CAZy family	pfam domain	Fibrobacteria								Chitinivibronia		Cellulolytic bacteria (35 genomes ¹)	Bacterial average across 2038 genomes	
		Fibrobacteraceae					Fibromonadaceae			<i>C. alkali-philus</i> AChT1	MC_77		Avg	Avg
		<i>F. succinogenes</i> S85	AD_80	AD_111	SR_36	AD_312	IN01_31	IN01_221	IN01_307					
Non-catalytic CBMs associated with:														
Cellulases														
CBM4	pfam02018	8.3	10.9	9.6	8.6	12.2	5.9	6.7	13.3	12.5	15.4	5.6	0.5	5.3
CBM30	-	0.0	0.0	0.0	1.7	0.0	5.9	6.7	0.0	0.0	0.0	0.6	0.1	1.2
CBM51	pfam14498	5.0	4.4	5.8	3.5	6.1	11.8	13.3	6.7	0.0	0.0	0.4	0.4	4.3
Subtotals (%)		13.3	15.2	15.4	13.8	18.4	23.5	26.7	20.0	12.5	15.4	6.6	1.0	
Chitinases														
CBM50	-	8.3	8.7	7.7	8.6	6.1	35.3	20.0	20.0	12.5	30.8	14.6	39.8	73.4
Hemicellulases (debranching and oligosaccharide-degrading enzymes)														
CBM11	pfam03425	6.7	8.7	7.7	6.9	8.2	23.5	26.7	26.7	25.0	0.0	0.1	0.0	0.6
CBM13	-	0.0	0.0	0.0	0.0	4.1	0.0	0.0	6.7	0.0	0.0	3.7	1.6	10.4
CBM22	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.4	6.2	0.4	4.0
CBM32	-	1.7	0.0	0.0	1.7	0.0	5.9	6.7	6.7	37.5	0.0	3.2	3.4	17.8
CBM35	-	23.3	17.4	19.2	27.6	16.3	0.0	0.0	6.7	0.0	0.0	3.7	0.7	7.3
CBM61	-	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.6	0.4	5.3
CBM67	-	1.7	2.2	1.9	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.8	1.2	8.1
Subtotals (%)		35.0	28.3	28.8	37.9	28.6	29.4	33.3	46.7	62.5	15.4	19.3	7.6	
Cellulases/hemicellulases														
CBM6	pfam03422	41.7	45.7	46.2	37.9	44.9	5.9	6.7	6.7	0.0	15.4	4.9	0.7	8.0
CBM9	pfam02018	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.7	5.0	0.9	8.1
Subtotals (%)		41.7	45.7	46.2	37.9	44.9	5.9	6.7	6.7	0.0	23.1	9.9	1.6	
Others														
CBM48	pfam02922	1.7	2.2	1.9	1.7	2.0	5.9	13.3	6.7	12.5	7.7	7.5	22.3	63.6
CBM66	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.7	1.1	1.0	7.4
Subtotals (%)		1.7	2.2	1.9	1.7	2.0	5.9	13.3	6.7	12.5	15.4	8.6	24.1	
Total estimated CBM		60	46	52	58	49	17	15	15	8	13			
Total estimated genes		2871	2754	3008	2906	2344	3391	3347	2773	2304	2321			
CBM to estimated genes (%)		2.1	1.7	1.7	2.0	2.1	0.5	0.4	0.5	0.3	0.6	1.4²	0.3³	

¹Based on Koeck et al. (2014)

²Average of 52 CBMs to total genes across 35 cellulolytic bacteria genomes

³Average of 52 CBMs to total genes across 2038 bacterial species

Table 5.4: Inventory of accessory attachment genes for polymer hydrolysis identified in the Fibrobacteres genomes, organised by functional category.

CAZy family	Function ID	Fibrobacteria								Chitinivibrionia		Cellulolytic bacteria (35 genomes ¹)	Bacterial average across 3454 sp. ²
		Fibrobacteraceae					Fibromonadaceae			<i>C. alkaliphilus</i> AChtl	MC_77		
		<i>F. succinogenes</i> S85	AD_80	AD_111	SR_36	AD_312	IN01_31	IN01_221	IN01_307			Avg	Avg
Other cellulose-binding proteins													
Fib_succ_major, corresponding to TIGR02145	pfam09603	54	65	72	96	17	119	116	70	0	41	0.0	0.1
fibro-slime protein, corresponding to TIGR02148	pfam07691	6	5	5	6	4	0	0	4	0	2	0.3	0.2
Type IV pilin genes													
Type IV pilin N-term methylation site PilA	pfam13544	3	3	2	3	3	4	2	2	6	2	2.2	2.7
type IV pilus assembly protein PilB	K02652	1	0	1	1	2	1	1	1	2	3	0.9	0.4
type IV pilus assembly protein PilC	K02653	1	1	1	1	1	1	1	1	2	1	0.8	0.4
leader peptidase (prepilin peptidase) / N-methyltransferase [EC:3.4.23.43 2.1.1.-]	K02654	1	1	1	1	1	1	1	1	1	1	1.1	0.5
type IV pilus assembly protein PilE	K02655	0	0	0	0	0	0	0	0	2	0	0.1	0.2
type IV pilus assembly protein PilM	K02662	1	1	1	1	1	1	1	1	1	1	0.7	0.3
type IV pilus assembly protein PilN	K02663	1	1	1	1	1	1	1	1	1	1	0.4	0.3
type IV pilus assembly protein PilO	K02664	1	1	1	1	1	1	1	1	1	1	0.3	0.2
type IV pilus assembly protein PilQ	K02666	2	2	2	2	2	1	1	1	1	1	0.1	0.2
two-component system, NtrC family, response regulator PilR	K02667	0	0	0	0	1	0	0	0	0	0	0.1	0.1
twitching motility protein PilT	K02669	3	2	3	2	3	2	2	2	3	2	0.9	0.5

¹Based on Koeck et al. (2014)

²Average across 3454 bacterial genomes on IMG

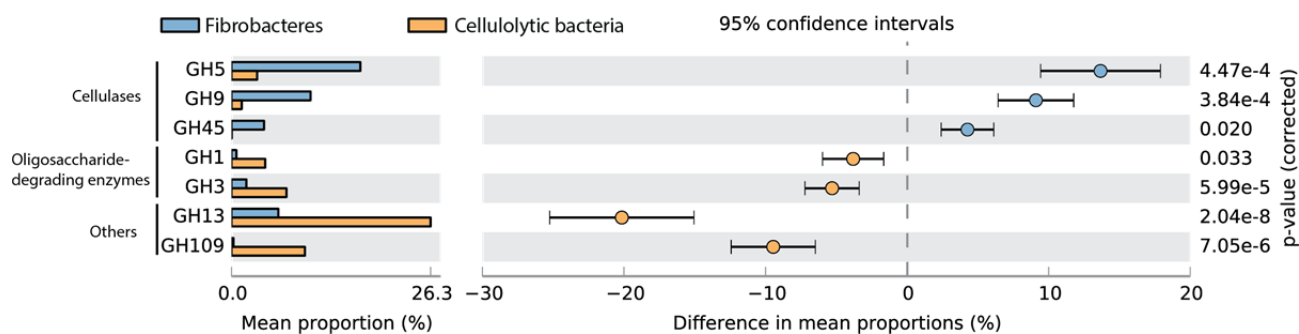


Figure 5.2: Glycoside hydrolase families with a significant difference in mean proportions $\geq 1\%$ between Fibrobacteres and other cellulolytic bacteria and a $p \leq 0.05$. Statistical significance was assessed using Welch's t-test with Bonferroni multiple test correction.

5.3.3.2 Fermentative metabolism and respiration

We expected that fermentation of sugars resulting from polymer hydrolysis would be the primary metabolism in the Fibrobacteres based on the obligate fermentative phenotype of *F. succinogenes* (Suen et al., 2011) and *C. alkaliphilus* (Sorokin et al., 2014). Metabolic reconstruction indicates that all Fibrobacteres genomes have the potential to utilise glucose via the Embden-Meyerhof pathway (EMP) and pentose phosphate pathway (PPP), but not via the Entner-Doudoroff pathway which is absent (**Figure 5.3**). It has previously been noted that *F. succinogenes* and *C. alkaliphilus* are unable to grow on xylan as a sole carbon source which suggests that they use their xylanases simply to expose cellulose and chitin respectively rather than using the resulting xylose as a growth substrate (Suen et al., 2011; Sorokin et al., 2014).

In that context, all Fibrobacteres lack the genes encoding a xylose permease and xylose interconversions via xylulose to xylulose-5-P which could then be processed via the PPP (**Figure 5.3**), suggesting the inability to use xylose is a phylum-level trait. The ability to use chitin hydrolysis products appears to be limited to the *Chitinivibrionia* genomes. All investigated Fibrobacteres should be able to perform the initial hydrolysis of insoluble chitin to smaller soluble oligosaccharides via GH18, which can be imported into the periplasm via TonB-dependent transporters (**Figure 5.3**). However, either GH19 (*C. alkaliphilus* only) or GH20 (MC_77 only) are required to hydrolyse the soluble oligosaccharides into N-acetylglucosamine (GlcNAc) dimers or trimers, which can then be converted into fructose-6-P and enter the EMP or PPP pathways (Sorokin et al., 2014). For all studied genomes, the end products of the EMP pathway, phosphoenolpyruvate and pyruvate, can then enter the tricarboxylic acid (TCA) cycle or the latter can be metabolised to formate, acetate or ethanol. All 10 genomes encode incomplete TCA cycles as they lack succinyl-CoA-synthase as previously noted for both *F. succinogenes* (Suen et al., 2011) and *C. alkaliphilus* (Sorokin et al., 2014). All Fibrobacteria also lack 2-oxoglutarate synthase and the two *Chitinivibrionia* representatives lack succinate dehydrogenase suggesting succinate and

fumarate are end products of the reductive arm of the TCA cycle for these classes, respectively (**Figure 5.3**). Succinate is a major fermentative end products of *F. succinogenes* (Suen et al., 2011), but fumarate does not accumulate as a fermentation product of *C. alkaliphilus* (Sorokin et al., 2014).

All of the investigated *Fibrobacteraceae* and *Chitinivibrionia* genomes lack major components of the electron transport chain (ETC) and are incapable of growth via respiration, which is consistent with previous reports that their characterised representatives are obligate anaerobes (Suen et al., 2011; Sorokin et al., 2014). By contrast, the *Fibromonadaceae* genomes encode an ETC comprising complexes I and II, cytochrome bd and an ATP synthase, which should be able to perform some form of electron-transport linked phosphorylation (**Figure 5.3** and Appendix D: **Table S5.3**). The cytochrome bd complex in other bacteria functions under low oxygen conditions (Borisov et al., 2011), which is consistent with the termite hindgut habitat from which the *Fibromonadaceae* genomes were obtained. Due to its small size, the termite hindgut is only anoxic in the central region and has microoxic peripheries (Brune et al., 1995). To investigate the origins of the *Fibromonadaceae* ETC, we inferred phylogenetic trees from the most conserved components (bd complex), which indicate that the common ancestor of the family had an ETC which is distantly related to other phyla and unlikely to be the result of a recent lateral transfer (Appendix D: **Figure S5.3**). Other lineages within the Fibrobacteres, currently lacking genomic representation (**Figure 5.1B**), may also have ETCs, which if present, will help to shed light on the ancestry of respiration in this phylum. All Fibrobacteres genomes, with the exception of MC_77, encode enzymes to counter oxidative stress including thioredoxin reductase and superoxide dismutase, but not catalase (Appendix D: **Table S5.3**). The apparent absence of antioxidant enzymes in MC_77 may be an artefact of the lower estimated completeness (73.3%) of this genome.

5.3.3.3 Nitrogen and ammonia metabolism

Lignocellulosic biomass is nitrogen limited and a poor source of amino acids, vitamins and their precursors (Brune, 2014). Metabolic reconstruction revealed a sporadic distribution of core nitrogen fixing genes (*nifH*, *nifD* and *nifK*) amongst the Fibrobacteres representatives (**Figure 5.1**, **Figure 5.3** and Appendix D: **Table S5.3**), suggesting a history of gain and loss by lateral gene transfer as previously noted more generally for nitrogen fixation (Boucher et al., 2003). We created phylogenetic trees for NifD and NifK and infer that the genes encoding these proteins were independently acquired relatively recently in the *Fibrobacteraceae*, *Fibromonadaceae* and *Chitinivibrionia* from different Firmicutes donors (Appendix D: **Figure S5.4**). Genes immediately flanking the *nif* genes were conserved in each family supporting lateral acquisition (Appendix D:

Figure S5.5). Our data are therefore not consistent with the idea of an early acquisition of nitrogen-fixing genes in the Fibrobacteres (Suen et al., 2011), but rather suggest a patchy history of recent gain and loss in habitats where nitrogen-fixing genes are present in numerous other community members providing the opportunity for lateral transfer (Warnecke et al., 2007; Brulc et al., 2009; He et al., 2013). Whether the *nif* genes are functionally active is debatable as *F. succinogenes*, which contains only four *nif* genes (3 core; *nifH,D,K*), has not been shown to be capable of nitrogen fixation (Suen et al., 2011). If any of the Fibrobacteres are capable of nitrogen fixation, they have amongst the lowest recorded number of subunits (3 to 9) for an active nitrogenase (Wang et al., 2013). By contrast, all members of the Fibrobacteres have ammonia uptake and assimilation genes (Appendix D: **Table S5.3**) which supply their nitrogen requirements (Matheron et al., 1999, Suen et al., 2011, He et al., 2013). All ten of the Fibrobacteres genomes have the potential to synthesise most of their own amino acids and cofactors (**Figure 5.3** and Appendix D: **Table S5.3**), including the gut symbionts, suggesting that they are not dependent on other organisms or host diet for most of their nutritional requirements.

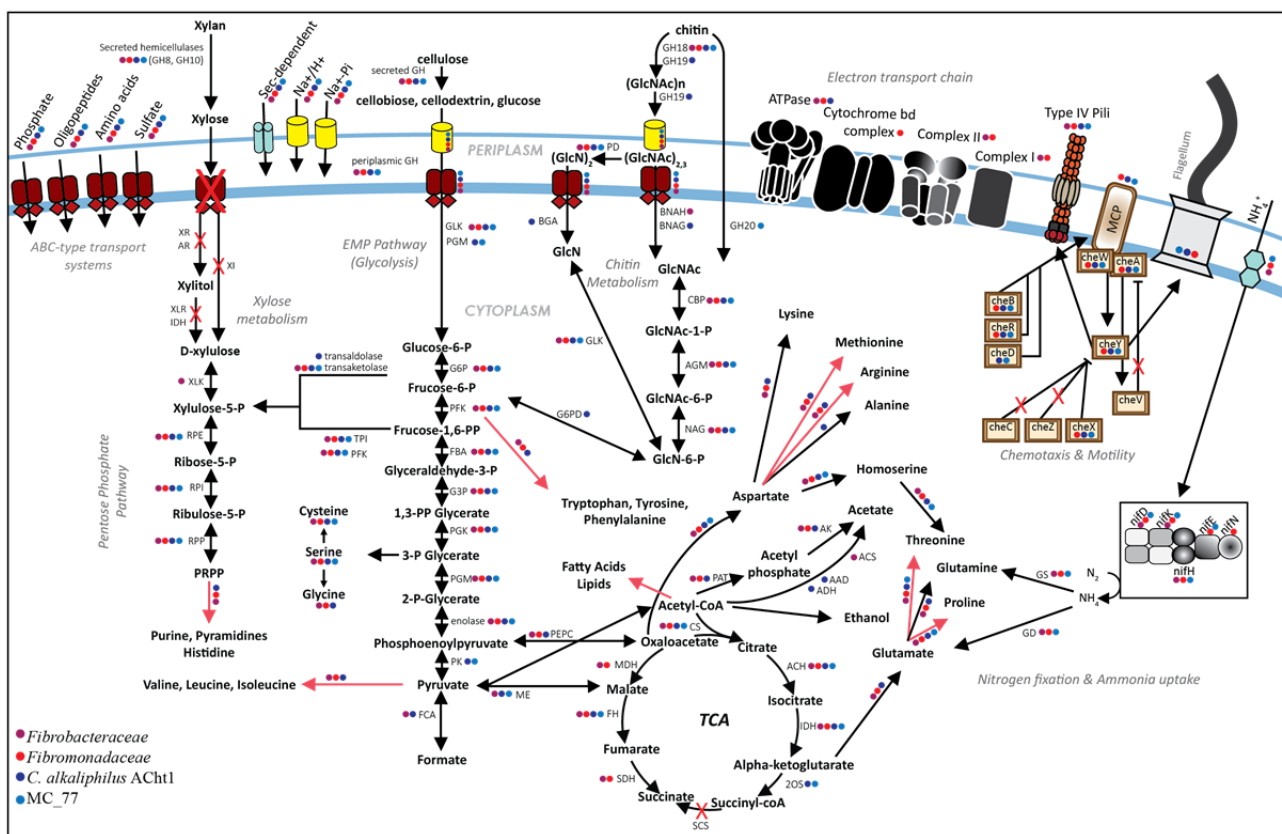


Figure 5.3: Composite metabolic reconstruction of members of the phylum Fibrobacteres. Presence of genes and pathways in a given lineage is indicated by coloured dots (legend at lower left). Steps in metabolic pathways absent in all investigated Fibrobacteres genomes are indicated by red crosses. Multistep reactions are shown by red arrows. Abbreviations are described in Appendix D: Table S5.3.

5.3.3.4 *Motility and chemotaxis*

Fibrobacteres have been defined as non-motile bacteria based on their only characterised representative genus, *Fibrobacter* (Ransom-Jones et al., 2012; Jewell et al., 2013). However, far from being a phylum-level trait, all investigated members of the *Fibromonadaceae* and *Chitinivibrionia* encode numerous flagellar and associated chemotaxis genes (**Figure 5.3**, Appendix D: **Figure S5.6** and **Table S5.3**), which is consistent with the direct observation of a polar flagellum in *C. alkaliphilus* (Sorokin et al., 2014). Methyl-accepting chemotaxis proteins were notably more abundant in the *Fibromonadaceae* and MC_77 genomes than in *C. alkaliphilus* (Appendix D: **Figure S5.6**) despite the closer phylogenetic relationship of MC_77 to *C. alkaliphilus*. This may reflect habitat differences since *Fibromonadaceae* and MC_77 reside in termite guts which have complex chemical milieus and steep chemical gradients likely requiring motile microorganisms to respond to a wider range of environmental cues than *C. alkaliphilus* in a hypersaline soda lake. Putative sensory hydrogenases were identified in members of both the *Fibromonadaceae* and MC_77 (Appendix D: **Figure S5.7**), which are hypothesised to allow these bacteria to orient themselves to steep hydrogen gradients present in the termite gut (Warnecke et al., 2007). The absence of flagella and chemotaxis previously reported for *F. succinogenes* (Suen et al., 2011) appears to be a family-level trait in the *Fibrobacteraceae* (**Figure 5.3**, Appendix D: **Figure S5.8** and **Table S5.3**). Phylogenetic analysis of several core flagellar genes (Liu and Ochman, 2007) suggests that motility was vertically inherited from a common Fibrobacteres ancestor and subsequently lost in the *Fibrobacteraceae* lineage (Appendix D: **Figure S5.8**). Since most members of this family have adapted to life in the herbivore gut, flagella-enabled chemotaxis and motility may have been no longer required due to an abundance of degradable substrates and mixing of contents provided by the host animal. Further genomic representation of the phylum will be required to determine if other lineages within the Fibrobacteres have similarly lost motility genes.

5.4 Conclusion

In this study, we have substantially expanded the phylogenomic representation of the Fibrobacteres and TG3 lineages by obtaining eight draft genomes of environmental populations from termite guts, anaerobic cellulose-fed digester and a sheep rumen. We propose that TG3 should be amalgamated with the Fibrobacteres phylum because the two lineages are robustly monophyletic in concatenated marker gene trees, and because they share a number of key traits. These include polymer hydrolysis which appears to be a unifying feature of the phylum, reflected by environmental distribution in habitats in which polymer hydrolysis plays a major role. As with *F. succinogenes*, all Fibrobacteres representatives have xylanases, but lack the genes necessary to metabolise xylan degradation

products for energy transduction. In contrast to previous suppositions largely based on characteristics of the genus *Fibrobacter*, we infer that not all members of the Fibrobacteres are strictly anaerobic as some have respiratory chains, and most appear to be motile. Members of the family *Fibromonadaceae* have low oxygen bd cytochromes allowing them to respire in microaerophilic conditions, and flagella-mediated motility is inferred to be an ancestral trait in the phylum having being lost from the family *Fibrobacteraceae*. Nitrogen fixing genes are sporadically distributed across the phylum and appear to have been obtained by multiple independent lateral transfers, whereas salvaging of fixed nitrogen from ammonia is inferred to be a more general trait. The eight population genomes described in the present study form an improved basis for further investigations into the Fibrobacteres phylum.

Description of *Candidatus Fibromonas termitidis*

Fibromonas termitidis (Fi.bro.mo'nas L. fem. n. *fibra*, fiber or filament in plants or animals; Gr. fem. n. *monas*, a unit, monad. ter.- mi'ti.dis. L. n. *tarmes*, *tarmit-* (L.L.var. *termes*, *termit-*) worm that eats wood; M.L. adj. *termitidis* pertaining to the termite). Not cultivated. Inferred to be Gram-negative, motile, containing an electron transport chain, and able to use cellulose as a primary growth substrate. Represented by population genome IN01_221 (acc. no. LIUG00000000) obtained from metagenomes of whole gut samples of the higher termite, *Microcerotermes* (acc. no. KJ907817).

Description of *Fibromonadaceae* (fam. nov.)

The description is the same as for the genus *Fibromonas*; *-aceae* ending to denote a family. Type genus: *Candidatus Fibromonas*

Description of *Fibromonadales* (ord. nov.)

The description is the same as for the genus *Fibromonas*; *-ales* ending to denote an order. Type family: *Fibromonadaceae* fam. nov.

Emended description of the phylum Fibrobacteres Garrity and Holt 2012

The phylum Fibrobacteres is a deep-branching lineage of the Bacteria. On the basis of comparative sequence analysis of isolate and environmental genomes, the phylum comprises at least two classes; *Fibrobacteria* and Chitinivibrionia, and three orders; *Fibrobacterales*, *Fibromonadales* and *Chitinivibrionales*. Gram-negative, polymer-degrading bacteria.

5.5 Acknowledgements

We thank the reviewers, Garret Suen and Jared Leadbetter, for their detailed and constructive comments, and Serene Low, Margaret Butler and Nicola Angel for help with preparing samples for Illumina sequencing. This research was supported by a Queensland Smart State Co-investment Fund grant awarded to PH and UQ strategic funding to the Australian Centre for Ecogenomics. NAR and IV were supported by UQ Research Scholarships. PH and DHP were supported by a Discovery Outstanding Researcher Award (DORA) and Australian Laureate Fellowship from the Australian Research Council (DP120103498 and FL150100038). GT was supported by a UQ VC Research Focused Fellowship.

5.6 References

- Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., and Nielsen, P.H. (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology* **31**: 533-538.
- Allgaier, M., Reddy, A., Park, J.I., Ivanova, N., D'haeseleer, P., Lowry, S. et al. (2010) Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS One* **5**: e8812.
- Beier, S., and Bertilsson, S. (2013) Bacterial chitin degradation—mechanisms and ecophysiological strategies. *Frontiers in Microbiology* **4**.
- Berlemont, R., and Martiny, A.C. (2013) Phylogenetic distribution of potential cellulases in bacteria. *Applied and Environmental Microbiology* **79**: 1545-1554.
- Borisov, V.B., Gennis, R.B., Hemp, J., and Verkhovsky, M.I. (2011) The cytochrome bd respiratory oxygen reductases. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1807**: 1398-1413.
- Boucher, Y., Douady, C.J., Papke, R.T., Walsh, D.A., Boudreau, M.E.R., Nesbø, C.L. et al. (2003) Lateral gene transfer and the origins of prokaryotic groups. *Annual review of genetics* **37**: 283-328.
- Brulc, J.M., Antonopoulos, D.A., Miller, M.E.B., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A. et al. (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences* **106**: 1948-1953.
- Brune, A. (2014) Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology* **12**: 168-180.
- Brune, A., Emerson, D., and Breznak, J.A. (1995) The termite gut microflora as an oxygen sink: microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. *Applied Environmental Microbiology* **61**: 2681-2687.
- Dai, X., Zhu, Y., Luo, Y., Song, L., Liu, D., Liu, L. et al. (2012) Metagenomic insights into the fibrolytic microbiome in yak rumen. *PLoS One* **7**: e40430.
- Dassa, B., Borovok, I., Ruimy-Israeli, V., Lamed, R., Flint, H.J., Duncan, S.H. et al. (2014) Rumen cellulosomes: divergent fiber-degrading strategies revealed by comparative genome-wide analysis of six ruminococcal strains. *PLoS One* **9**: e99221.
- Dupont, C.L., Rusch, D.B., Yooseph, S., Lombardo, M.-J., Alexander Richter, R., Valas, R. et al. (2012) Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *The ISME Journal* **6**: 1186-1199.
- Giltner, C.L., Nguyen, Y., and Burrows, L.L. (2012) Type IV pilin proteins: versatile molecular modules. *Microbiology and Molecular Biology Reviews* **76**: 740-772.
- Gupta, R.S. (2004) The phylogeny and signature sequences characteristics of Fibrobacteres, Chlorobi, and Bacteroidetes. *Critical Reviews in Microbiology* **30**: 123-143.

- He, S., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R.H. et al. (2013) Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS One* **8**: e61126.
- Hongoh, Y., Deevong, P., Hattori, S., Inoue, T., Noda, S., Noparatnaraporn, N. et al. (2006) Phylogenetic diversity, localization, and cell morphologies of members of the candidate phylum TG3 and a subphylum in the phylum Fibrobacteres, recently discovered bacterial groups dominant in termite guts. *Applied and Environmental Microbiology* **72**: 6780-6788.
- Hongoh, Y., Deevong, P., Inoue, T., Moriya, S., Trakulnaleamsai, S., Ohkuma, M. et al. (2005) Intra- and interspecific comparisons of bacterial diversity and community structure support coevolution of gut microbiota and termite host. *Applied and Environmental Microbiology* **71**: 6590-6599.
- Hungate, R.E. (1950) The anaerobic mesophilic cellulolytic bacteria. *Bacteriological Reviews* **14**: 1-49.
- Imelfort, M., Parks, D., Woodcroft, B.J., Dennis, P., Hugenholtz, P., and Tyson, G.W. (2014) GroopM: an automated tool for the recovery of population genomes from related metagenomes. *PeerJ* **2**: e603.
- Iyo, A.H., and Forsberg, C.W. (1996) Endoglucanase G from *Fibrobacter succinogenes* S85 belongs to a class of enzymes characterized by a basic C-terminal domain. *Canadian Journal of Microbiology* **42**: 934-943.
- Jewell, K.A., Scott, J.J., Adams, S.M., and Suen, G. (2013) A phylogenetic analysis of the phylum *Fibrobacteres*. *Systematic and Applied Microbiology* **36**: 376-382.
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C. et al. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236-1240.
- Koeck, D.E., Pechtl, A., Zverlov, V.V., and Schwarz, W.H. (2014) Genomics of cellulolytic bacteria. *Current Opinion in Biotechnology* **29**: 171-183.
- Konstantinidis, K.T., and Tiedje, J.M. (2005) Towards a genome-based taxonomy for prokaryotes. *Journal of Bacteriology* **187**: 6258-6264.
- Krieg, N.R., Parte, A., Ludwig, W., Whitman, W.B., Hedlund, B.P., Paster, B.J. et al. (2011) The Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes. In *Bergey's Manual of Systematic Bacteriology*: Springer, pp. 596 -603.
- LeClerc, G.R., Buchan, A., Maurer, J., Moran, M.A., and Hollibaugh, J.T. (2007) Comparison of chitinolytic enzymes from an alkaline, hypersaline lake and an estuary. *Environmental Microbiology* **9**: 197-205.
- Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:13033997*.
- Li, Y., Sun, H., Ma, X., Lu, A., Lux, R., Zusman, D., and Shi, W. (2003) Extracellular polysaccharides mediate pilus retraction during social motility of *Myxococcus xanthus*. *Proceedings of the National Academy of Sciences* **100**: 5443-5448.

- Lightfield, J., Fram, N.R., and Ely, B. (2011) Across bacterial phyla, distantly-related genomes with similar genomic GC content have similar patterns of amino acid usage. *PLoS One* **6**: e17677.
- Liu, R., and Ochman, H. (2007) Stepwise formation of the bacterial flagellar system. *Proceedings of the National Academy of Sciences* **104**: 7116-7121.
- Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**: 490-495.
- Ludwig, W., and Klenk, H.-P. (2001) Overview: a phylogenetic backbone and taxonomic framework for procaryotic systematics. In *Bergey's Manual of Systematic Bacteriology*: Springer, pp. 49-65.
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Buchner, A. et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Research* **32**: 1363-1371.
- Malburg, L., Iyo, A.H., and Forsberg, C.W. (1996) A novel family 9 endoglucanase gene (celD), whose product cleaves substrates mainly to glucose, and its adjacent upstream homolog (celE) from *Fibrobacter succinogenes* S85. *Applied and Environmental Microbiology* **62**: 898-906.
- Markowitz, V.M., Chen, I.-M.A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M. et al. (2014) IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic acids research* **42**: 568-573.
- Matheron, C., Delort, A.-M., Gaudet, G., Liptaj, T., and Forano, E. (1999) Interactions between carbon and nitrogen metabolism in *Fibrobacter succinogenes* S85: a ¹H and ¹³C nuclear magnetic resonance and enzymatic study. *Applied and Environmental Microbiology* **65**: 1941-1948.
- Mikaelyan, A., Köhler, T., Lampert, N., Rohland, J., Boga, H., Meuser, K., and Brune, A. (2015) Classifying the bacterial gut microbiota of termites and cockroaches: a curated phylogenetic reference database (DictDb). *Systematic and Applied Microbiology* **38**: 472-482.
- Miron, J., Ben-Ghedalia, D., and Morrison, M. (2001) Invited review: adhesion mechanisms of rumen cellulolytic bacteria. *Journal of Dairy Science* **84**: 1294-1309.
- Montgomery, L., and Macy, J.M. (1982) Characterization of rat cecum cellulolytic bacteria. *Applied and Environmental Microbiology* **44**: 1435-1443.
- Montgomery, L., Flesher, B., and Stahl, D. (1988) Transfer of *Bacteroides succinogenes* (Hungate) to *Fibrobacter succinogenes* gen. nov. as *Fibrobacter succinogenes* comb. nov. and description of *Fibrobacter intestinalis* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* **38**: 430-435.
- Morrison, M., and Miron, J. (2000) Adhesion to cellulose by *Ruminococcus albus*: a combination of cellulosomes and Pil-proteins? *FEMS microbiology letters* **185**: 109-115.
- Morrison, M., Pope, P.B., Denman, S.E., and McSweeney, C.S. (2009) Plant biomass degradation by gut microbiomes: more of the same or something new? *Current Opinion in Biotechnology* **20**: 358-363.
- Nawrocki, E.P., Kolbe, D.L., and Eddy, S.R. (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**: 1335-1337.

- Parks, D.H., Tyson, G.W., Hugenholtz, P., and Beiko, R.G. (2014) STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**: 3123-3124.
- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*: **25**:1043–1055.
- Pegden, R.S., Larson, M.A., Grant, R.J., and Morrison, M. (1998) Adherence of the gram-positive bacterium *Ruminococcus albus* to cellulose and identification of a novel form of cellulose-binding protein which belongs to the Pil family of proteins. *Journal of Bacteriology* **180**: 5921-5927.
- Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**: 785-786.
- Price, M.N., Dehal, P.S., and Arkin, A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution* **26**: 1641-1650.
- Qi, M., Jun, H.-S., and Forsberg, C.W. (2007) Characterization and synergistic interactions of *Fibrobacter succinogenes* glycoside hydrolases. *Applied and Environmental Microbiology* **73**: 6098-6105.
- Qi, M., Jun, H.-S., and Forsberg, C.W. (2008) Cel9D, an atypical 1, 4- β -d-glucan glucohydrolase from *Fibrobacter succinogenes*: characteristics, catalytic residues, and synergistic interactions with other cellulases. *Journal of Bacteriology* **190**: 1976-1984.
- Qi, M., Nelson, K.E., Daugherty, S.C., Nelson, W.C., Hance, I.R., Morrison, M., and Forsberg, C.W. (2005) Novel molecular features of the fibrolytic intestinal bacterium *Fibrobacter intestinalis* not shared with *Fibrobacter succinogenes* as determined by suppressive subtractive hybridization. *Journal of Bacteriology* **187**: 3739-3751.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. et al. (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**: 590-596.
- Rahman, N.A., Parks, D.H., Willner, D.L., Engelbrektson, A.L., Goffredi, S.K., Warnecke, F. et al. (2015) A molecular survey of Australian and North American termite genera indicates that vertical inheritance is the primary force shaping termite gut microbiomes. *Microbiome* **3**: 5.
- Ransom-Jones, E., Jones, D.L., McCarthy, A.J., and McDonald, J.E. (2012) The Fibrobacteres: an important phylum of cellulose-degrading bacteria. *Microbial Ecology* **63**: 267-281.
- Ransom-Jones, E., Jones, D.L., Edwards, A., and McDonald, J.E. (2014) Distribution and diversity of members of the bacterial phylum Fibrobacteres in environments where cellulose degradation occurs. *Systematic and Applied Microbiology* **37**: 502-509.
- Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30.14**: 2068-2069.
- Sethi, A., and Scharf, M.E. (2013) Biofuels: fungal, bacterial and insect degraders of lignocellulose. In: *Encyclopedia of Life Science (eLS)*: Chichester: John Wiley and Sons, Ltd.
- Soo, R.M., Skennerton, C.T., Sekiguchi, Y., Imelfort, M., Paech, S.J., Dennis, P.G. et al. (2014) An expanded genomic representation of the phylum Cyanobacteria. *Genome Biology and Evolution* **6**: 1031-1045.

- Sorokin, D., Tourova, T., Sukhacheva, M., Mardanov, A., and Ravin, N. (2012) Bacterial chitin utilisation at extremely haloalkaline conditions. *Extremophiles* **16**: 883-894.
- Sorokin, D.Y., Gumerov, V.M., Rakitin, A.L., Beletsky, A.V., Damsté, J., Muyzer, G. et al. (2014) Genome analysis of *Chitinivibrio alkaliphilus* gen. nov., sp. nov., a novel extremely haloalkaliphilic anaerobic chitinolytic bacterium from the candidate phylum Termite Group 3. *Environmental Microbiology* **16**: 1549-1565.
- Spain, A.M., Forsberg, C.W., and Krumholz, L.R. (2010) Phylum XVIII. Fibrobacteres Garrity and Holt 2001. In *Bergey's Manual® of Systematic Bacteriology*: Springer, pp. 737-746.
- Standley, K. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability.(outlines version 7). *Molecular Biology and Evolution* **30**: 772-780.
- Suen, G., Weimer, P.J., Stevenson, D.M., Aylward, F.O., Boyum, J., Deneke, J. et al. (2011) The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. *PloS One* **6**: e18814.
- Tokuda, G., Tsuboi, Y., Kihara, K., Saitou, S., Moriya, S., Lo, N., and Kikuchi, J. (2014) Metabolomic profiling of ¹³C-labelled cellulose digestion in a lower termite: insights into gut symbiont function. *Proceedings of the Royal Society of London B: Biological Sciences* **281**: 20140990.
- Toyoda, A., Iio, W., Mitsumori, M., and Minato, H. (2009) Isolation and identification of cellulose-binding proteins from sheep rumen contents. *Applied and Environmental Microbiology* **75**: 1667-1673.
- Vanwonterghem, I., Jensen, P.D., Dennis, P.G., Hugenholtz, P., Rabaey, K., and Tyson, G.W. (2014) Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. *The ISME Journal* **8**: 2015-2028.
- Wang, L., Zhang, L., Liu, Z., Zhao, D., Liu, X., Zhang, B. et al. (2013) A minimal nitrogen fixation gene cluster from *Paenibacillus* sp. WLY78 enables expression of active nitrogenase in *Escherichia coli*. *PLoS Genetics* **9**: e1003865.
- Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T. et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**: 560-565.
- Wilson, D.B. (2009) Evidence for a novel mechanism of microbial cellulose degradation. *Cellulose* **16**: 723-727.
- Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **40**: 445-451.

Chapter 6 Conclusion and future directions

6.1 Conclusion

Termites (order Isoptera) are a group of important eusocial insects that have been the subject of extensive basic and applied research. They thrive on recalcitrant wood lignocellulose as their main food source while some have utilised other nutritionally poor biomass like soil, grass and herbivore dung. Their ability to survive on these substrates is through a mutualistic symbiosis between termite host and gut microorganisms. In lower termites, this nutritional interaction involves a tripartite symbiosis of the host, protists and prokaryotes, but in higher termites the association has been evolutionarily reduced to host and prokaryotes. This has led to fundamental research in understanding termite physiology, phylogeny and gut microbiology. Termites as global pests have contributed major economic losses for damage to agriculture and structures. From an applied research perspective, this has led to the integration of pest management and biotechnological applications in studying termite biology (Scharf, 2015). Prior to the omics era, pioneering research using culture- and molecular SSU-based approaches have shaped our understanding of termite gut microbiology, yet further analyses are impeded due to limitations of these methods such as the inability to examine the overall functional potential of the termite gut. The advancement of sequencing capacity and computational power in the omics era has not only resulted in the great majority of termite gut microbiology studies being conducted in the last decade but has provided new views on the complex microbial gut symbiosis (Brune and Dietrich, 2015).

The focus of this thesis was to characterise the gut communities of primarily Australian termites via high throughput culture-independent approaches. This is aligned with the goal of answering broader ecological questions that include the effect of diet versus co-evolution, effect of changing diet on microbial community structure and function, the function of specific populations, and relative contributions of prokaryotic and eukaryotic symbionts to lignocellulose hydrolysis in lower termites. As little molecular data exist for most Australian termite fauna, we profiled the gut microbiomes of 66 Australian and American termites using SSU rRNA amplicon pyrosequencing (Chapter 2). To our knowledge, this represents the first culture-independent gut microbiome profiles of three higher termite genera (*Tenuirostritermes*, *Drepanotermes*, and *Gnathamitermes*) and two lower termite genera (*Marginitermes* and *Porotermes*). In addition to primary findings that indicate vertical inheritance is the major factor in structuring the termite gut microbiome, we also found that diet leaves an imprint influencing the relative abundance but not membership of gut communities. We hypothesised that changes in relative abundance can occur on shorter timescales as a form of

adaptation for dietary fluctuations. This hypothesis was tested in Chapter 3 through a series of feeding assays using the polyphagous Australian termite species, *Mastotermes darwiniensis* as the model organism. Changes in community composition of *Mastotermes* gut profiles were notable with compositionally different feedstocks (e.g. wood to grass) supporting our initial hypothesis. However, greater shifts appeared to be stress related, possibly due to a combination of dietary change and smaller colony size. Despite the altered composition changes, the gut functions were maintained with only small differences detected in corresponding gut protein profiles. In Chapter 4, we performed a gene-centric analysis of shotgun data comparing gut community functional capabilities of lower and higher termites. The similar gene abundance profiles between the four termite genera suggest a convergence for essential functions of the termite gut ecosystems, despite conspicuous differences in community composition. From the gut metagenomic datasets, we were able to recover four populations belonging to the Fibrobacteres phylum using a differential coverage binning approach. Fibrobacteres are minor constituents of most habitats but were strikingly abundant in the gut of wood-feeding higher termites. In Chapter 5, we performed comparative genomic analysis of these Fibrobacteres population genomes together with those acquired from cellulose-fed anaerobic digesters, rumen and soda lake. Metabolic reconstruction of these Fibrobacteres genomes provided insights into the metabolism of this underrepresented but important bacterial phylum in termite gut and other anaerobic ecosystems. In contrast to previous suppositions that Fibrobacteres are non-motile based on the only described genus *Fibrobacter*, our findings suggest that motility is an ancestral and widespread trait between members and has been lost in the family *Fibrobacteraceae*. Although polymer hydrolysis appears to be a prominent feature of members of Fibrobacteres, more representation should be investigated to confirm this inference.

6.2 Future directions

Community profiling using phylogenetic markers such as the SSU rRNA gene (16S and 18S rRNA genes) have been widely applied to characterise the 1 μ l-scale termite gut ecosystem (**Table 1.3**). It has not only provided useful insights into the density and diversity of the termite gut microbiota but also evidence for the coevolutionary patterns of host and their gut microorganisms (e.g. survey studies). The distinctive gut microbiota across the termite host phylogeny, to that of other insect gut communities, reflects their eusociality through proctodeal trophallaxis and coprophagy, ensuring that species are successfully and faithfully transferred from one generation to another. The termite gut microbiome is highly distinctive with a combination of core and accessory gut bacterial phyla among termite major groups (e.g. lower and higher termite). It is important to note that a dominant host signal in gut community composition does not imply that all component species are the product

of vertical inheritance which ultimately results in co-speciation of specific lineages. The differences in abundance of bacterial phyla between termite host groups are likely to fulfil different functional niches (Brune and Dietrich, 2015). As noted in Chapter 2, diet appears to be a secondary signal in shaping the termite gut microbiota of both lower and higher termites when considering both protist and prokaryotic communities. Diet however appears as a driving factor influencing the gut community structure when only bacterial communities, likely acquired from the environment, are taken into consideration (Leber and Balkwill, 1997; Colman et al., 2012; Dietrich et al., 2014; Tai et al., 2015). It is not surprising that such conclusions are drawn as most of the SSU rRNA surveys have extensively been applied to the identification of prokaryotic populations in the termite gut. SSU surveys have mainly focus on prokaryotic communities due to the extreme variation in eukaryotic SSU rRNA gene copy number in eukaryotes and primer mismatches (Amaral-Zettler et al., 2009). It is important that the entire community of the termite hindgut ecosystem is evaluated to capture the overall diversity. In addition, greater potential now lies in characterising the protist symbionts in lower termites using SSU rRNA surveys, provided that some requirements such as development of appropriate primers and analytical tools are met (Tai et al., 2013; Tai et al., 2015). Future SSU rRNA profiling efforts should be invested into the association of prokaryotic and flagellated protist populations to provide a detailed characterisation of the overall microbial diversity in this complex symbiosis expanding the understanding of current termite hindgut ecology. There is also a need for higher resolution phylogenetic analysis of candidate symbionts (Mikaelyan et al., 2015b), which cannot be attained through analysis of short reads, to provide concrete evidence for co-speciation (Mikaelyan et al., 2015a).

The failure to maintain social behaviour (e.g. social grooming) within a colony leads to the inability to sustain community stability that ultimately results in the breakdown of the gut community. It is worth noting that environmental conditions may also impact the termite gut community as observed through a shift in specific microbial populations over a period of seven days likely in response to laboratory conditions (Chapter 3). Similar feeding trials with the primary goal of discovering novel biocatalysts for second generation biofuel, focusing on different feedstocks of interest, have showed strong discrepancies in results between studies (Brune and Dietrich, 2015). Most of these studies have only focused on the “end of feeding assay” profiles without exploring underlying mechanisms such as temporal changes of gut profiles in relation to diet and physiological conditions of individuals. Hence, conclusions extrapolated from feeding experiments have to be considered with caution. Despite the bacterial communities exhibiting resistance to dietary perturbation, more feeding assays are required to address whether compositional changes is reflective of diet or simply stress. In the future, feeding experiments should be conducted at larger scale in terms of the number

of termite individuals and the size of enclosed space to minimise stress. Additionally, monitoring the colony fitness and survival rate would be useful in understanding the impact of perturbation on termite gut community stability.

There has been an inclining omic-based research towards understanding termite and gut symbiont biology, and their potential for biotechnological applications with the advancement of high-throughput sequencing technologies. Even after decades of research, the functional roles of individual microbial populations in the termite gut remains vague. The use of SSU rRNA-based approaches coupled with omics techniques has only outlined the major functional niches of these populations in the complex lul-hindgut environment. Although these techniques have provided identification of key functions in the termite gut ecosystem, SSU rRNA amplicon-based amplification introduces primer and PCR biases that lead to variations in community structure as observed between the pyrotag and mitag *Mastotermes* gut profiles (Chapter 3). Therefore, there is a need to move towards metagenomic community profiling to provide a less biased view of the gut microbial diversity. We also noted that from the amplicon-based profiles, the archaeal relative abundance was substantially higher and variable in a number of species (e.g. *Porotermes* and *Mastotermes*) than previously appreciated, likely due to primer biased. However, the higher proportion of archaea was also reflected in *Porotermes* metagenome-based profiles (Chapter 4) which confirms our initial findings. As we move beyond SSU rRNA-based identification of termite gut symbionts, a follow-up experiment to investigate the archaeal populations would involve overlaying metagenomics and metatranscriptomics, coupled with visualisation methods. By examining the expression of the gut populations and their spatial distributions, it would be possible to provide an understanding of the potential functionality of the high abundance of the archaeal community. Additionally, the 16S rRNA analysis has limited resolution that results in closely related sequences representing a single OTU at 97% similarity. The multiple and variable 16S rRNA copies do not provide breakdown of functionality which genomics-based studies allow. Moreover, shotgun metagenomics should be able to address the aforementioned on-going conflicting opinions as to whether the termite gut microbiota is reflective of cospeciation or dietary fluctuations without the biases inherent to primer and PCR amplification, providing a higher resolution profiling.

Since termites and their gut symbionts are identified as a potential resource for industrial applications, omic-based studies have mainly focused on the aspects of lignocellulose digestion. It will be interesting to determine if the unexpected presence of low abundance ciliates in some higher termite genera are directly involved in lignocellulose digestion (Chapter 2). There are also many opportunities to explore other aspects of termite gut symbiosis such as identification of key players

in major functions in lower termites, for instance, the abundant OTUs that were identified in this thesis such as *Fusobacterium* and *Dysgonomonas*. With a range of binning tools that are readily available (Thomas et al., 2012), underrepresented populations continue to be explored. Some populations that are prevalent in termite guts but have very low abundance (e.g. candidate phylum ZB-3 and SR1) would be of interest for comparative studies. Comparative analysis of population genomes recovered from termite gut and other anaerobic environments, such as members of the bacterial phylum Fibrobacteres, would provide an evolutionary perspective in addition to the goal of reconstructing the tree of life. Prediction of functional and catalytic features of population genomes remains a challenge based on analysis of putative gene products alone (Brune, 2014). Furthermore, most of the termite gut microbiota lack cultured representatives due to the difficulty in culturing *in vitro*. There should be an adjunct goal in isolating key members of termite gut microbiota (Brune, 2014). Novel termite gut microbes have been successfully isolated using two strategies; (1) selective substrates (Wertz and Breznak, 2007; Paul et al., 2012) or (2) unconventional cultivation (Leadbetter et al., 1999; Geissinger et al., 2009). Together with information gathered from genomic data, a rational isolation strategy would provide potential in expanding termite gut microorganism representation (Brune, 2014). This would definitely be an invaluable asset for characterising a wider diversity of Fibrobacteres to understand the ecophysiological properties that are not apparent from genome analysis alone.

6.3 References

- Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., and Huse, S.M. (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS One* **4**: e6372.
- Brune, A. (2014) Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology* **12**: 168-180.
- Brune, A., and Dietrich, C. (2015) The gut microbiota of termites: Digesting the diversity in the light of ecology and evolution. *Annual Review of Microbiology* **69**.
- Colman, D.R., Toolson, E.C., and Takacs-Vesbach, C.D. (2012) Do diet and taxonomy influence insect gut bacterial communities? *Molecular Ecology* **21**: 5124-5137.
- Dietrich, C., Köhler, T., and Brune, A. (2014) The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Applied and Environmental Microbiology* **80**: 2261-2269.
- Geissinger, O., Herlemann, D.P., Mörschel, E., Maier, U.G., and Brune, A. (2009) The Ultramicrobacterium “*Elusimicrobium minutum*” gen. nov., sp. nov., the First Cultivated Representative of the Termite Group 1 Phylum. *Applied and Environmental Microbiology* **75**: 2831-2840.

- Leadbetter, J.R., Schmidt, T.M., Graber, J.R., and Breznak, J.A. (1999) Acetogenesis from H₂ plus CO₂ by spirochetes from termite guts. *Science* **283**: 686-689.
- Leber, T.M., and Balkwill, F.R. (1997) Zymography: A single-step staining method for quantitation of proteolytic activity on substrate gels. *Analytical Biochemistry* **249**: 24-28.
- Mikaelyan, A., Dietrich, C., Köhler, T., Poulsen, M., Sillam-Dussès, D., and Brune, A. (2015a) Diet is the primary determinant of bacterial community structure in the guts of higher termites. *Molecular Ecology* **24**: 5284–5295.
- Mikaelyan, A., Köhler, T., Lampert, N., Rohland, J., Boga, H., Meuser, K., and Brune, A. (2015b) Classifying the bacterial gut microbiota of termites and cockroaches: a curated phylogenetic reference database (DictDb). *Systematic and Applied Microbiology*.
- Paul, K., Nonoh, J.O., Mikulski, L., and Brune, A. (2012) “*Methanoplasmatales*,” *Thermoplasmatales*-related archaea in termite guts and other environments, are the seventh order of methanogens. *Applied and Environmental Microbiology* **78**: 8245-8253.
- Scharf, M.E. (2015) Omic research in termites: an overview and a roadmap. *Frontiers in Genetics* **6**.
- Tai, V., James, E.R., Perlman, S.J., and Keeling, P.J. (2013) Single-cell DNA barcoding using sequences from the small subunit rRNA and internal transcribed spacer region identifies new species of *Trichonympha* and *Trichomitopsis* from the hindgut of the termite *Zootermopsis angusticollis*. *PLoS One* **8**: e58728.
- Tai, V., James, E.R., Nalepa, C.A., Scheffrahn, R.H., Perlman, S.J., and Keeling, P.J. (2015) The role of host phylogeny varies in shaping microbial diversity in the hindguts of lower termites. *Applied and Environmental Microbiology* **81**: 1059-1070.
- Thomas, T., Gilbert, J., and Meyer, F. (2012) Metagenomics—a guide from sampling to data analysis. *Microbial Informatics and Experimentation* **2**: 1-12.
- Wertz, J.T., and Breznak, J.A. (2007) Physiological ecology of *Stenoxybacter acetivorans*, an obligate microaerophile in termite guts. *Applied and Environmental Microbiology* **73**: 6829-6841.

Appendix A: Supplementary figures and tables for Chapter 2

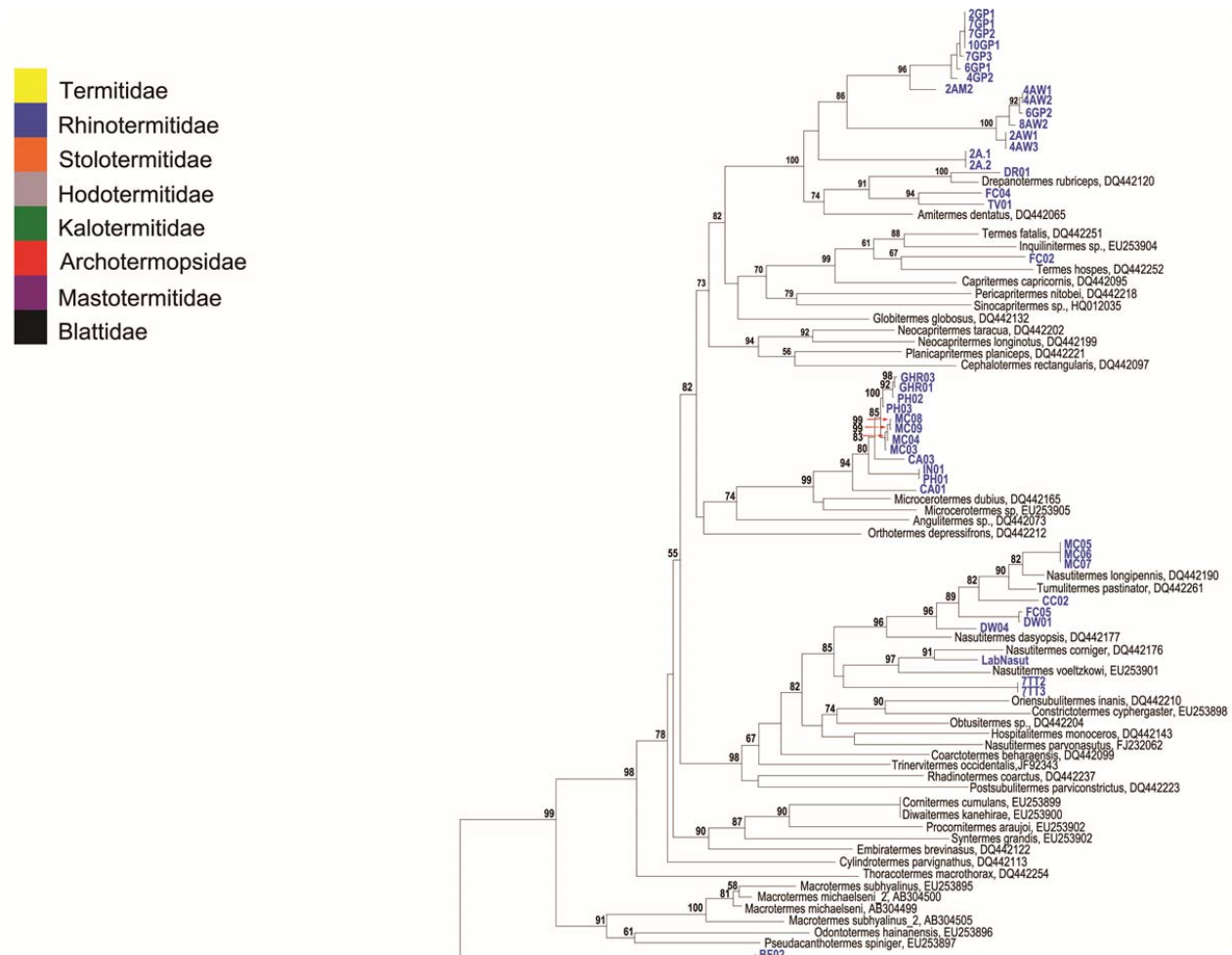


Figure S2.1: Maximum likelihood (FastTree) tree of aligned mitochondrial cytochrome oxidase (COII) genes from termite samples included in this study (in blue) and publicly available reference sequences. Family level affiliations are indicated by color according to the legend at left.

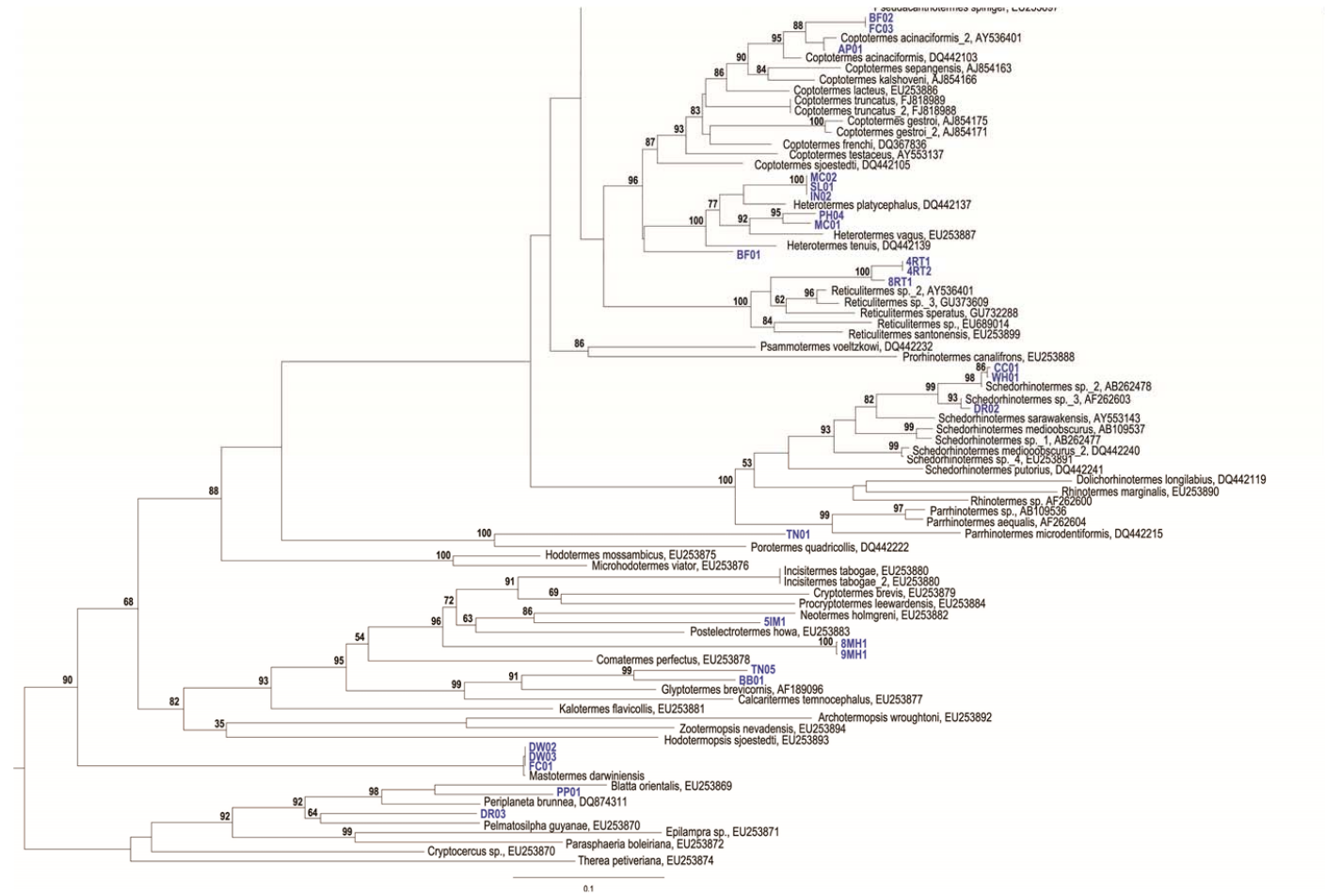


Figure S2.1: Continued

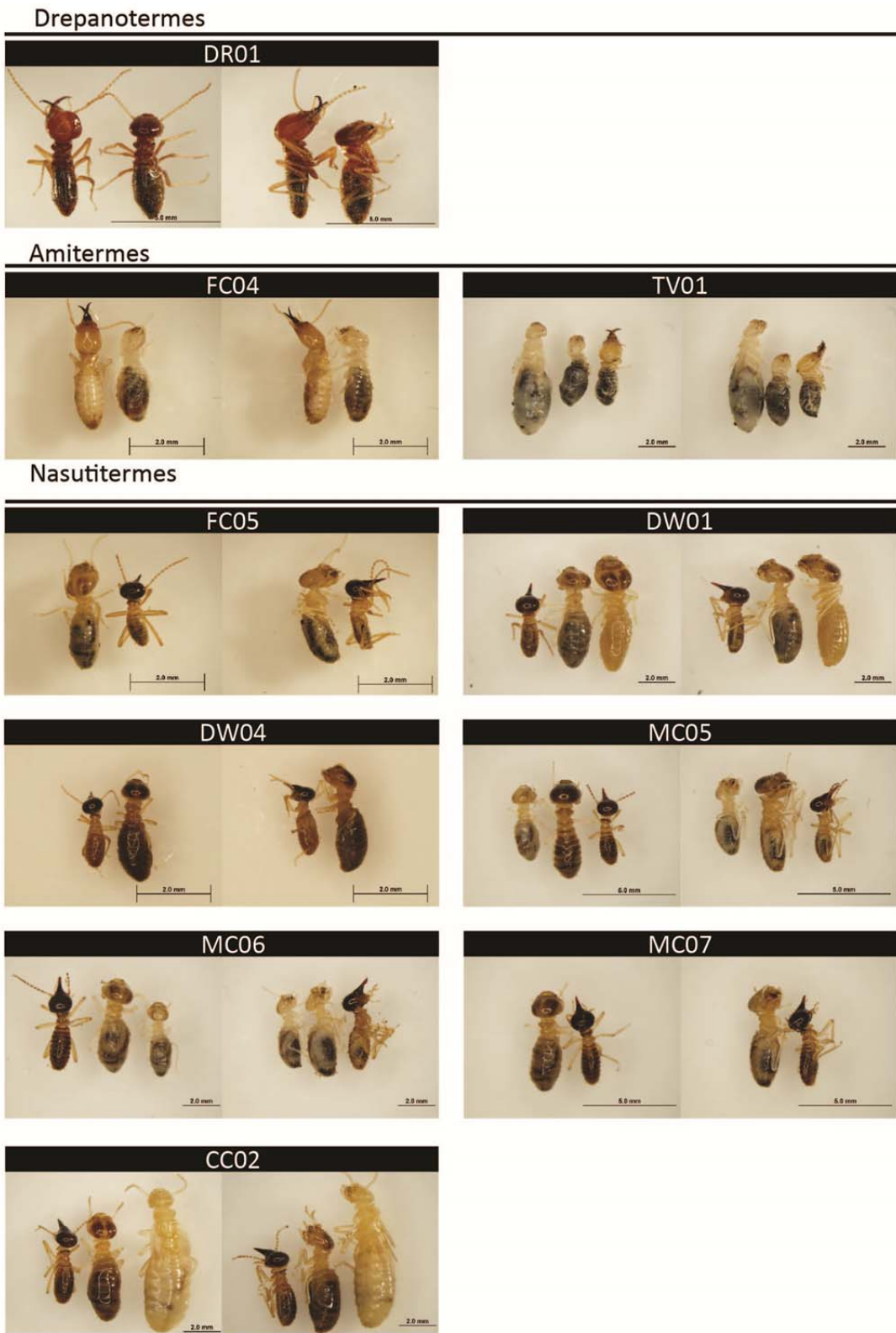


Figure S2.2: Soldier morphologies of several termite specimens collected in Australia.

Microcerotermes

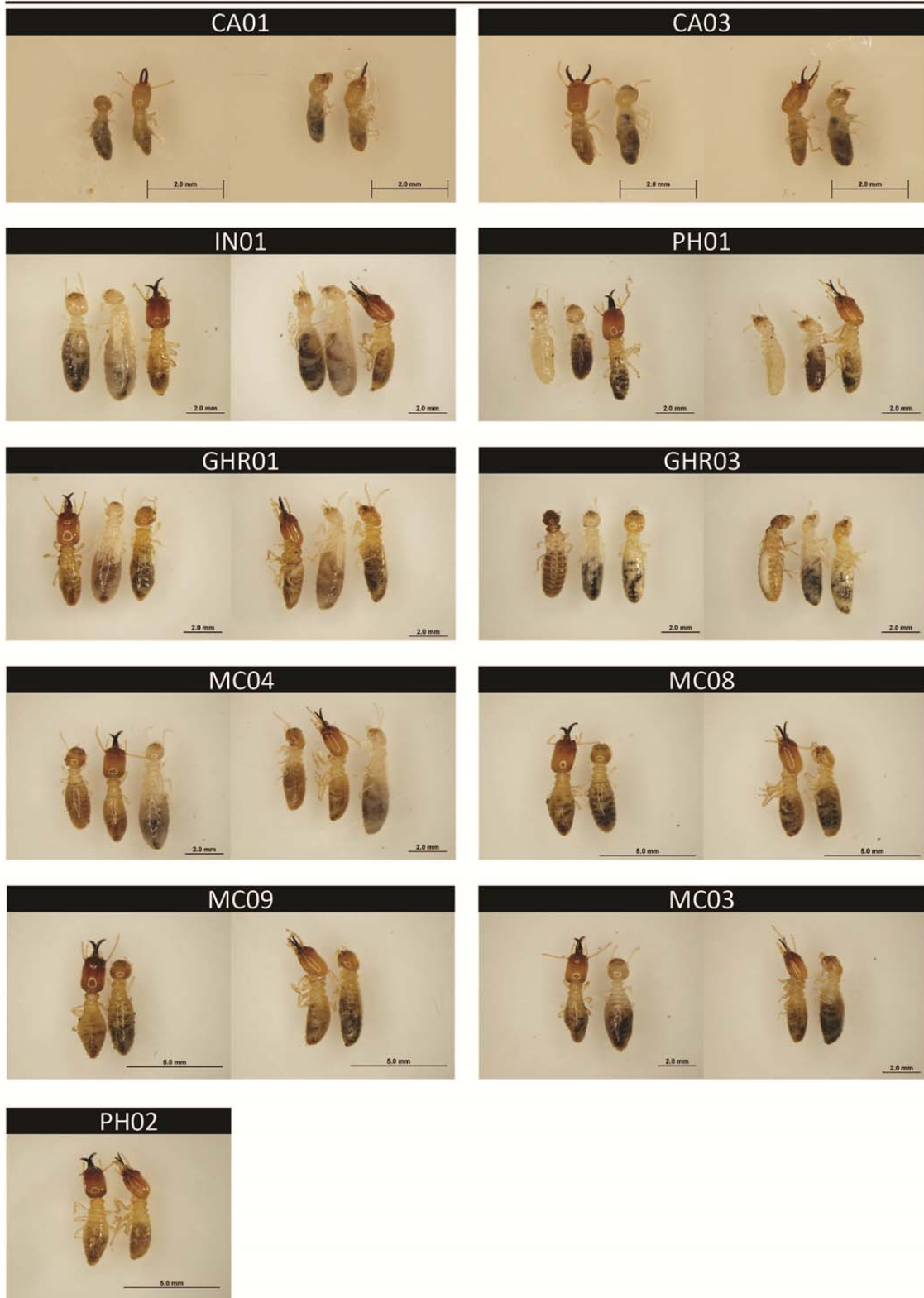
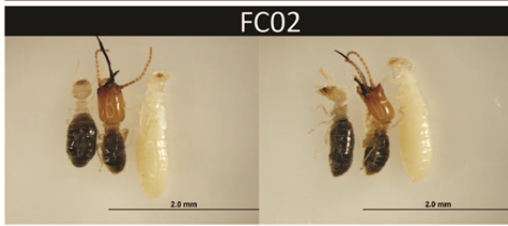
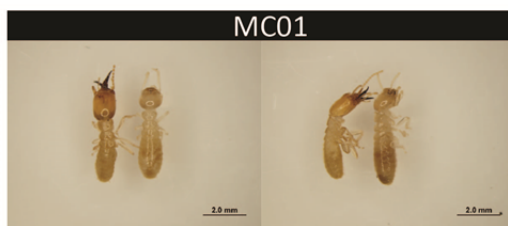
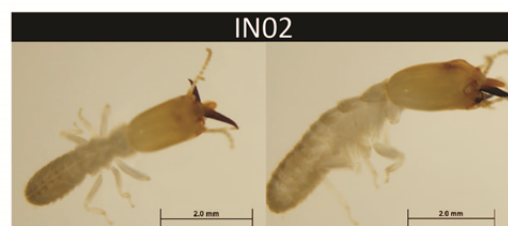
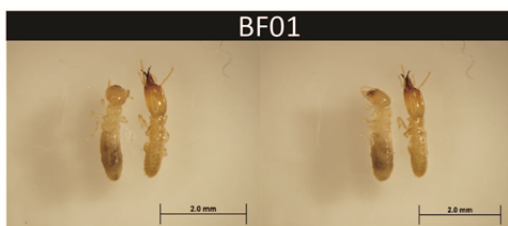
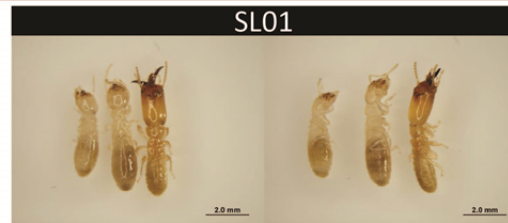
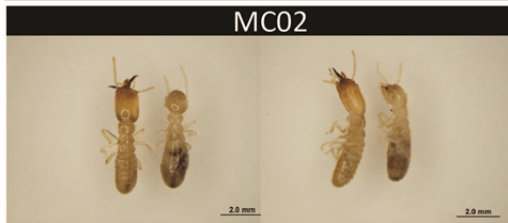


Figure S2.2: Continued

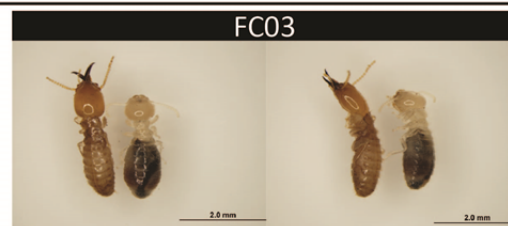
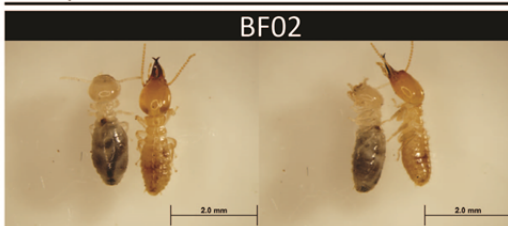
Macrognathotermes



Heterotermes



Coptotermes



Schedorhinotermes

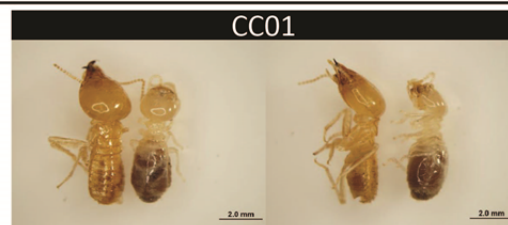
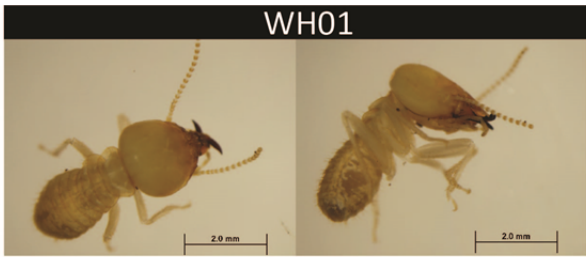
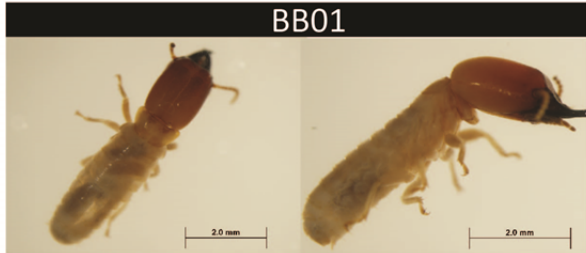


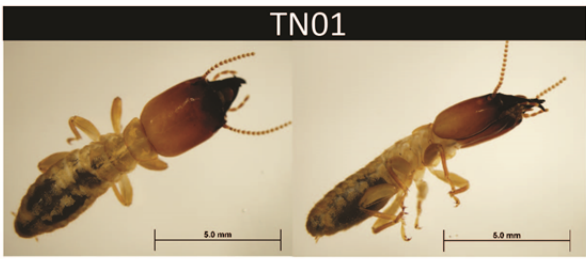
Figure S2.2: Continued



Glyptotermes



Porotermes



Mastotermes

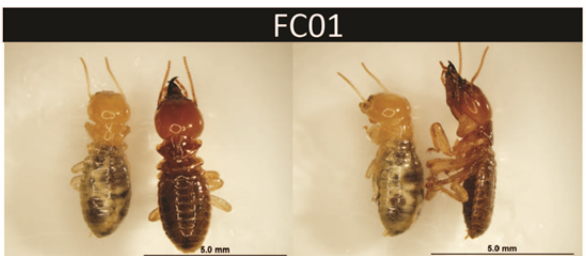
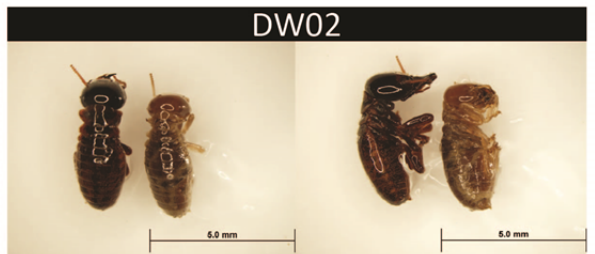
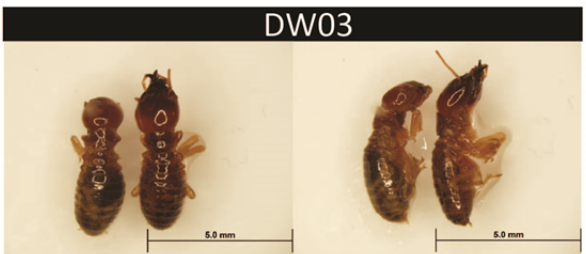


Figure S2.2: Continued

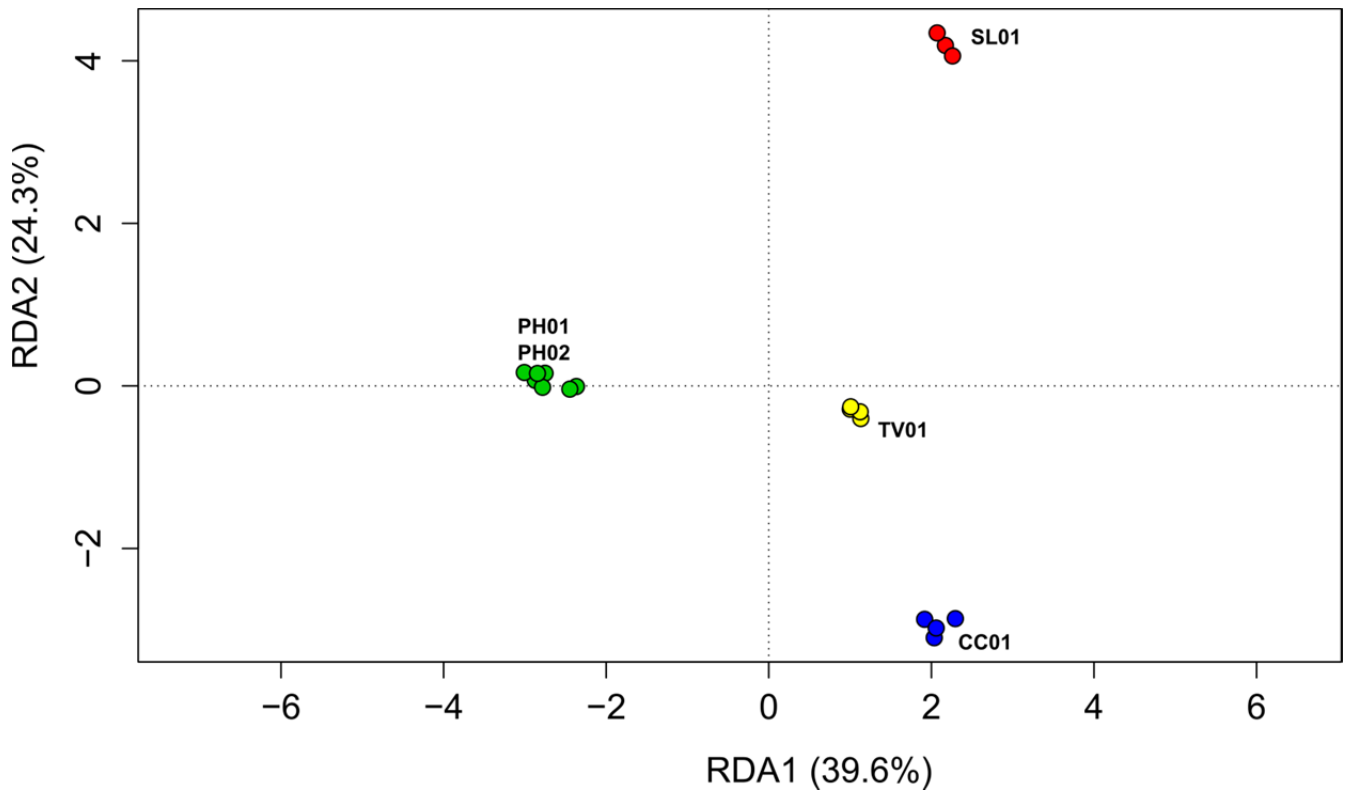


Figure S2.3: Redundancy analysis (RDA) plots of microbial profiles obtained from biological replicates of four termite genera. Differences were significantly less between biological replicates than between genera.

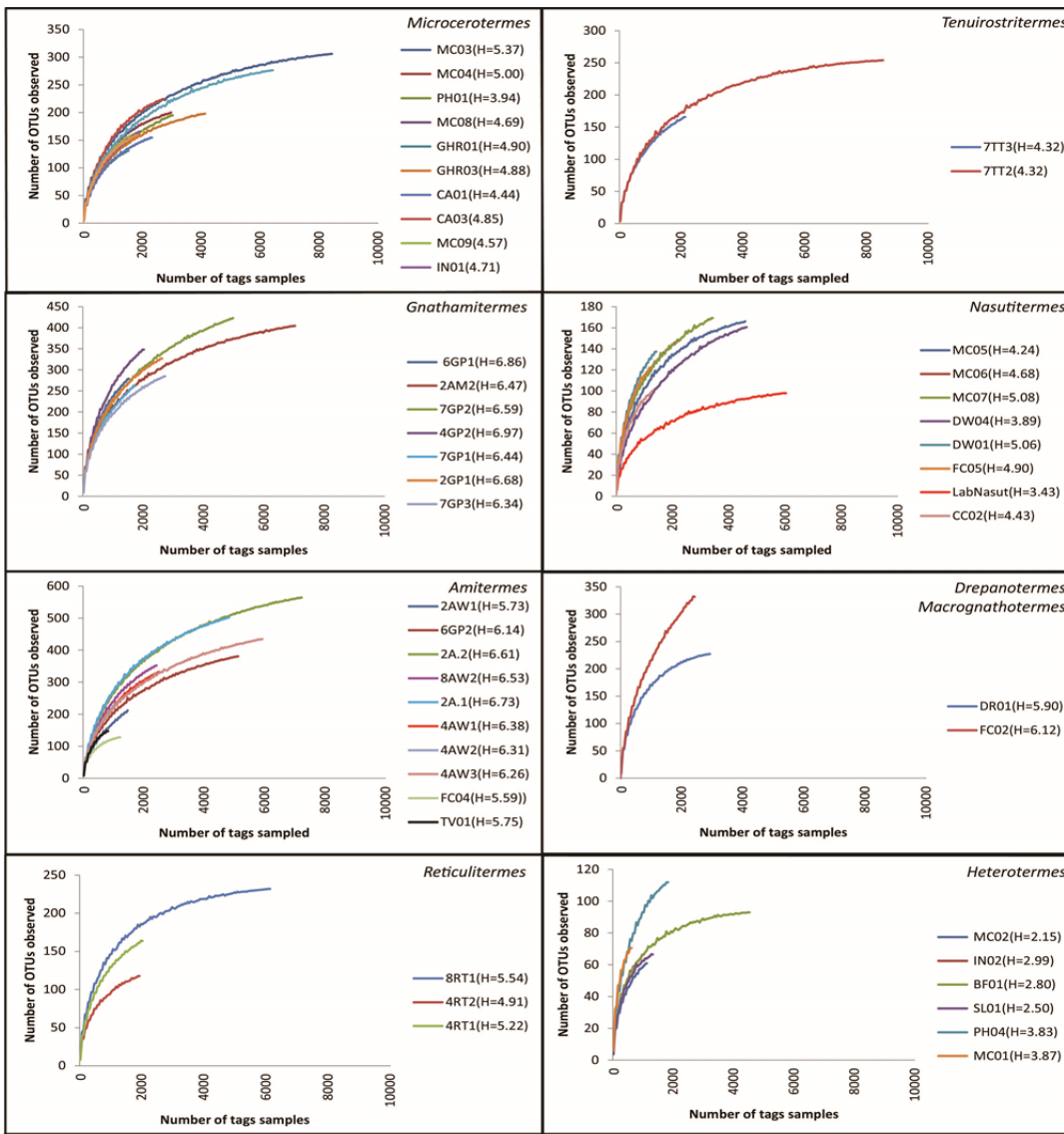


Figure S2.4: Rarefaction curves and associated Shannon diversity indices (H) of microbial profiles obtained for each of the 66 samples separated into different panels by termite genus affiliation.

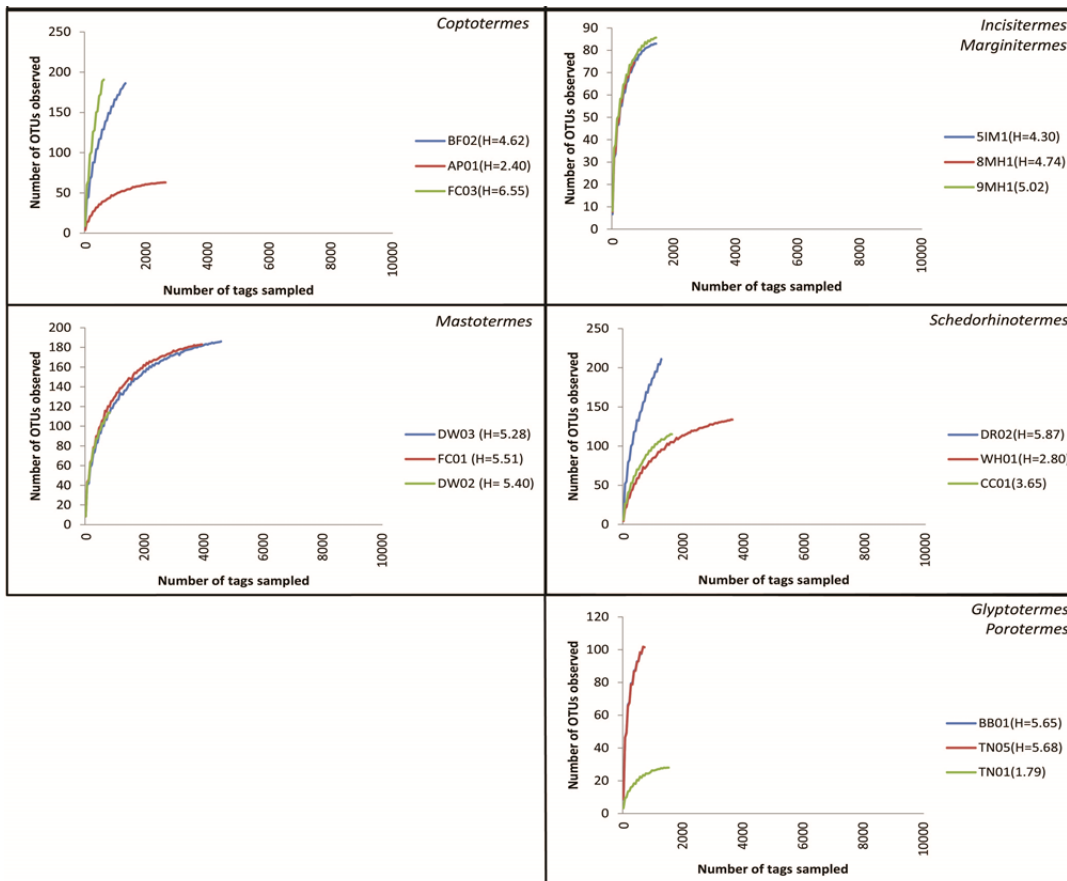


Figure S2.4: Continued

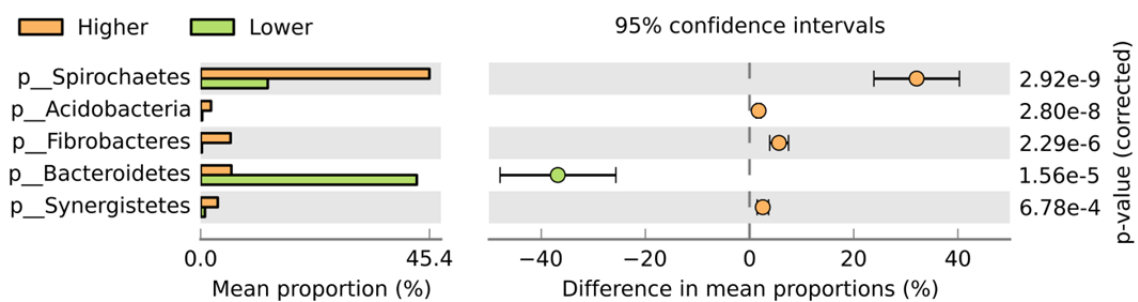


Figure S2.5: Core and accessory bacterial phyla with a significant difference in mean proportions $\geq 1\%$ between higher and lower termites and a p value ≤ 0.05 . Statistical significance was assessed using Welch's t-test with Šidák multiple test correction.

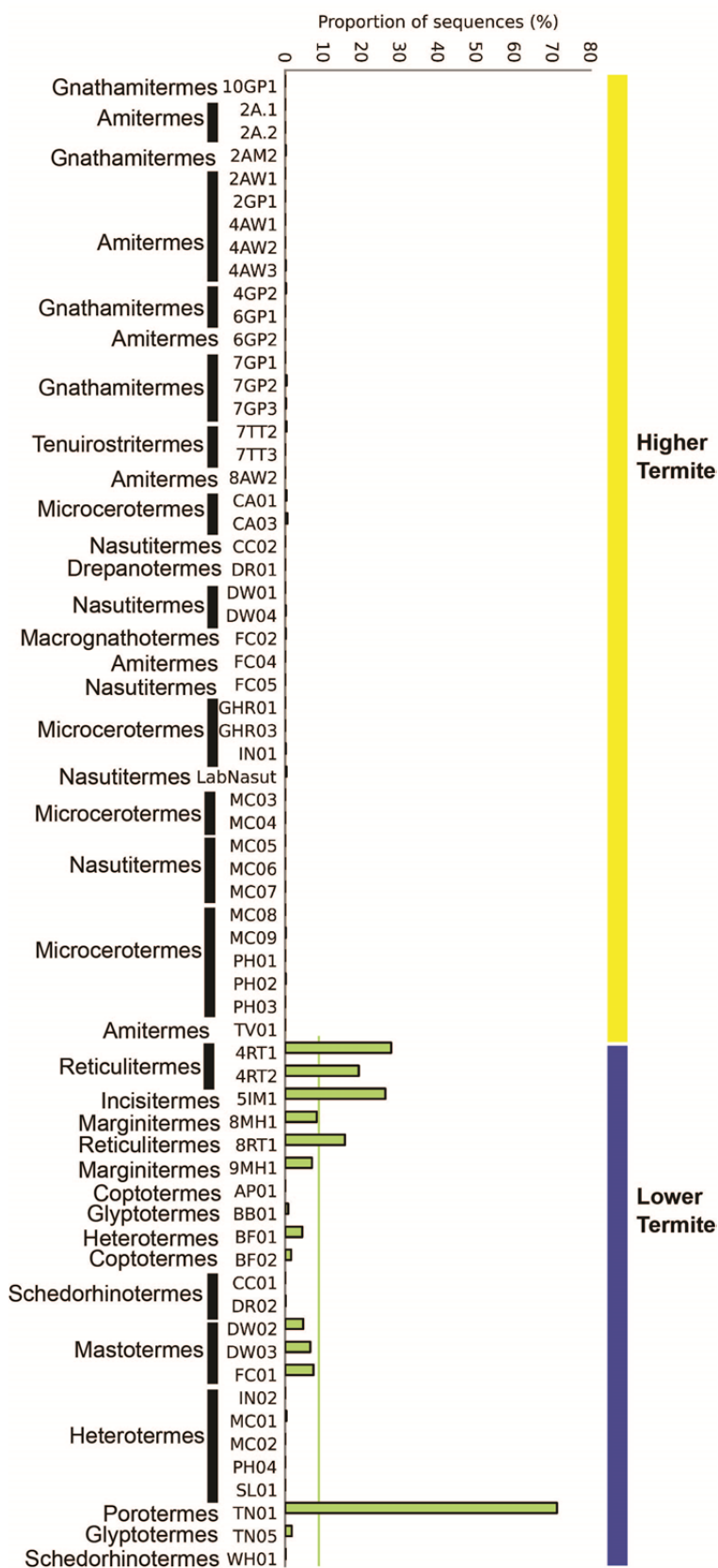


Figure S2.6: Relative proportion of Elusimicrobia across the higher and lower termite samples. Termite genus affiliations of the samples is shown to the left of the figure.

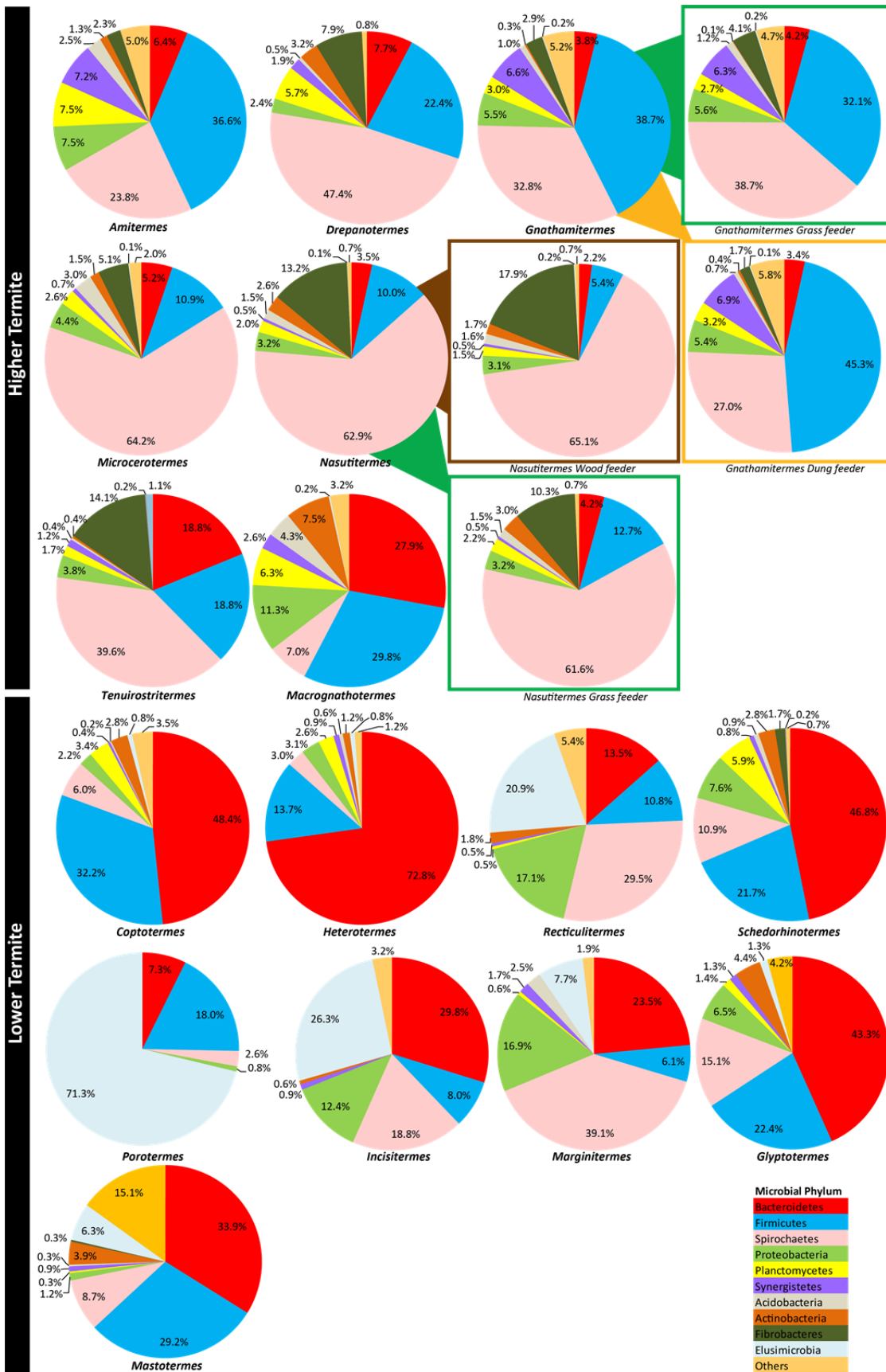


Figure S2.7: Average whole gut microbial community profiles of the 16 termite genera surveyed in this study. The profiles of the polyphagous termite genera *Gnathamitermes* and *Nasutitermes* are further divided by diet (in colored boxes).

Relative abundance



Macronathotermes		Coptotermes						Schedorhinotermes						Porotermes		Mastotermes				Consensus Lineage	#OTU ID			
FC02-803	FC02-927	AP01-803	AP01-926	FC03-803	FC03-926	BF02-803	BF02-926	CC01-803	CC01-926	DR02-803	DR02-926	WH01-803	WH01-926	TN01-803	TN01-926	FC01-803	FC01-926	DW03-803	DW03-926	DW02-803	DW02-926			
9.3	5.9			7.0	15.4	0.2	3.9	0.2		0.2													g__pGfC24	9
0.2				0.2																			g__pGfC25	1189
0.2				0.2																			g__pGfC26	1215
				0.2																			g__pGfC26	1013
0.1				0.1		0.1		0.1								1.1	1.5	0.1	3.3	3.9	4.8		g__Methanobacteriaceae	1091
		14.8	18.7			0.3	0.1	28.0	3.3			23.9	23.3	0.3	0.1								g__Methanobacterium	14
						0.1	0.1							44.1	54.8	1.3	4.5	0.1	4.0	3.1	3.0		g__Methanobacterium	643
						13.3	21.8	0.1				0.2	0.4										g__Methanorevibacter	2
								0.2															g__Methanorevibacter	662
								0.1															g__Methanorevibacter	4
								0.2															g__Methanorevibacter	621
								0.1															g__Methanorevibacter	1232
								0.1															g__Methanorevibacter	2226
									0.1														g__Methanorevibacter	730
														1.6	1.4								g__Methanorevibacter	1140
																							g__Methanorevibacter	1876
3.1	3.1			3.1	8.2		2.1	1.0	2.3	0.7	0.7	0.1	0.1			0.5	0.9	0.1	1.4	1.9	5.0		g__Methanorevibacter	23
3.1	3.1			1.4	3.5	0.1	0.9	0.3	0.4			0.3	0.2			2.9	5.6		7.0	3.6	6.1		g__Methanorevibacter	13
0.7	0.6			0.4	1.5		0.4									0.5	0.7						g__Methanorevibacter	76
0.5	0.2			0.1	0.5		0.1							0.3	0.5				0.1	0.2	1.5		g__Methanorevibacter	188
0.1				0.1																			g__Methanorevibacter	1821
0.1	0.1																						g__Methanorevibacter	1219
				0.1																			g__Methanorevibacter	1986
								0.1															g__Methanorevibacter	1117
								0.1															g__Methanorevibacter	2468
								0.1															g__Methanorevibacter	970
										0.1													g__Methanorevibacter	2316
											0.1												g__Methanorevibacter	2447
																							g__Methanorevibacter	1526
																0.1			0.1	0.1			g__Methanorevibacter	1199
																0.1	0.1		0.1	0.1			g__Methanorevibacter	1885
																					0.3		g__Methanorevibacter	1070
				0.1				0.1															g__Methanocella	2078
				0.1	0.1																		g__Methanospirillum	326
																	0.1	0.1					g__Methanospirillum	97
																		0.1	0.8	0.3	0.1		g__Methanospirillum	1813
0.5	0.6			0.6	1.6		0.4																g__Methanomicrococcus	41
0.1																							g__Methanosarcina	1193
8.7	6.3			5.4	9.1	0.2	2.3	9.3	19.2	0.2	0.1												g__Methanomassiliococcea	10
0.5	0.2			0.2	2.1		0.5																g__Methanomassiliococcea	580
0.2				0.0	0.0											0.2	0.5						g__Methanomassiliococcea	1633
0.2										0.2	0.2												g__Methanomassiliococcea	247
				0.2																			g__Methanomassiliococcea	2223
								0.2															g__Methanomassiliococcea	913
								0.2															g__Methanomassiliococcea	731
								0.2															g__Methanomassiliococcea	2239

Figure S2.9: Heatmap of archaeal OTUs generated with two primer pairs in whole gut samples of termites with $\geq 10\%$ archaeal relative abundance (Table 2). Each row represents a different OTU, and the abundance as a percentage of the total community is indicated by shading according to the legend. Termite family affiliations of each sample are indicated at the top the figure, respectively, and OTU phylogeny is indicated to the left (phylum) and right (mostly genus) of the figure.

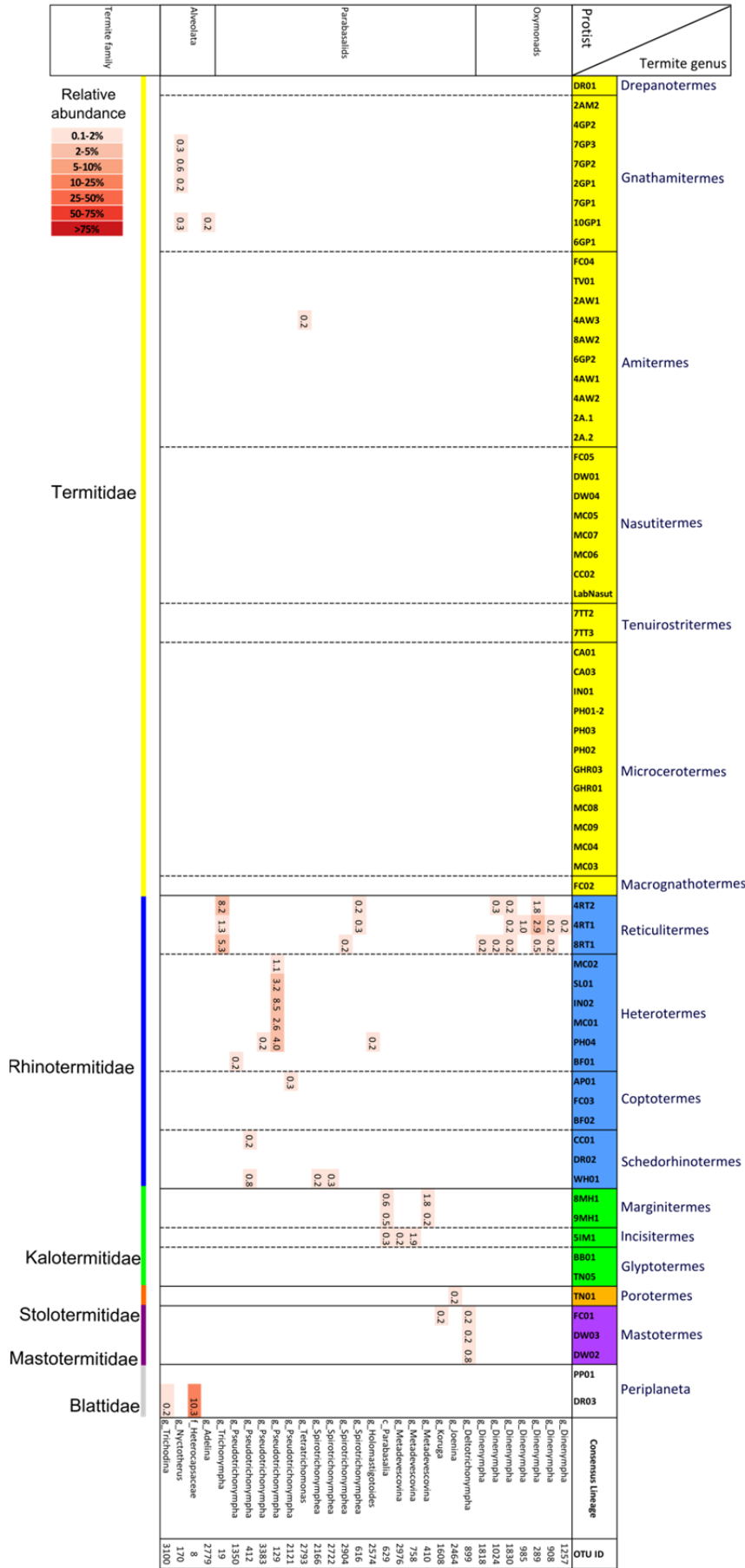


Figure S2.10: Heatmap of protist OTUs (97% seq id) across the 66 termite samples. Each row represents an OTU and each column a gut sample with relative abundance as a percentage of the total microbial community indicated by numbers and shading according to the legend. The termite genus and family for each sample is indicated at the top and bottom of the figure, respectively, and OTU phylogeny is indicated to the left (phylum) and right (mostly genus) of the figure.

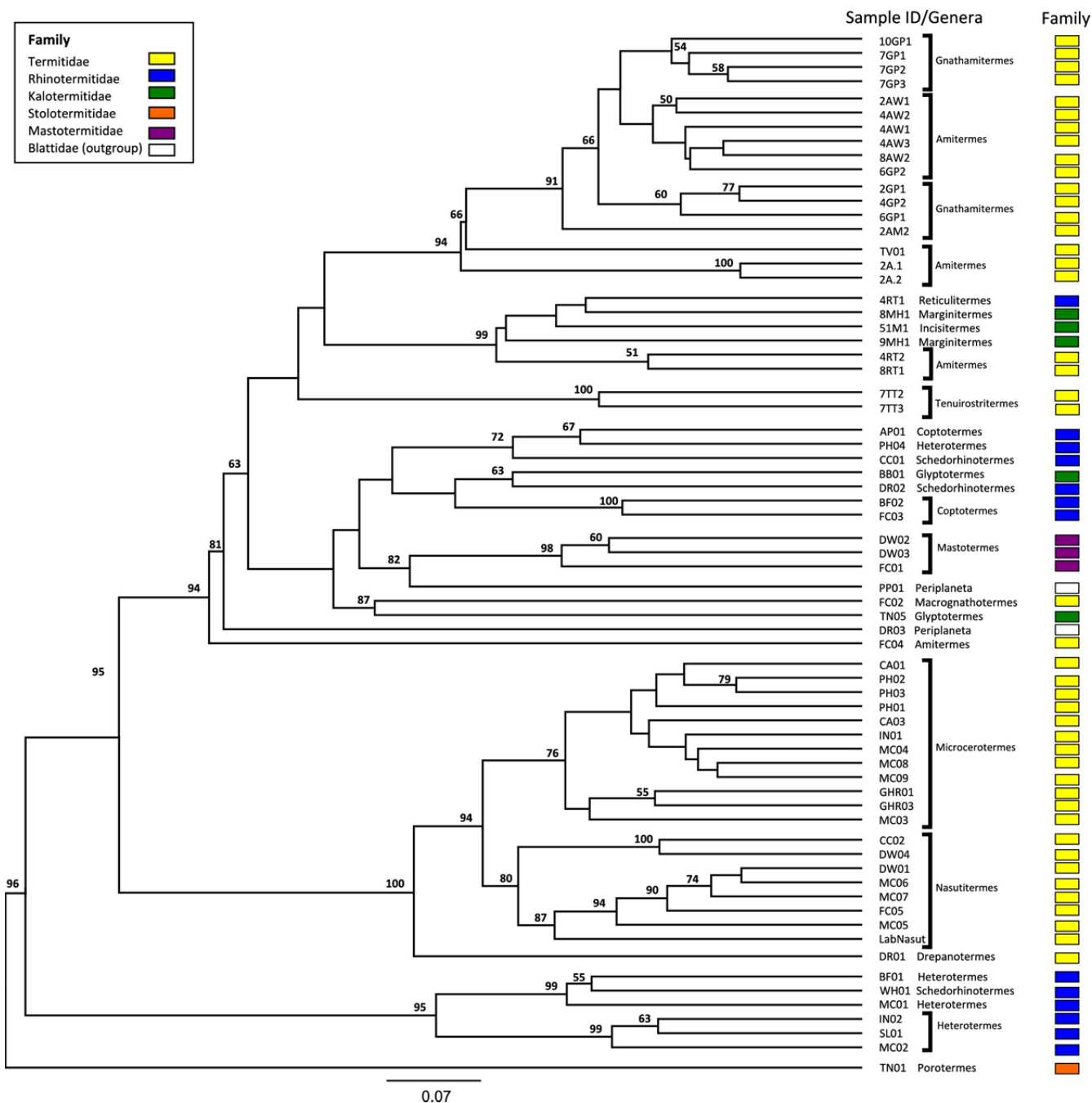


Figure S2.11: UPGMA tree of weighted (relative abundance taken into account) Soergel pairwise distances between bacterial profiles showing a drop in consistency with host phylogeny (particularly family level) relative to the unweighted analysis (Figure 4; Additional file 14: Table S4). The values on interior nodes represent jackknife support values ≥ 49 . Termite host affiliation (family) is indicated to the right of the tree.

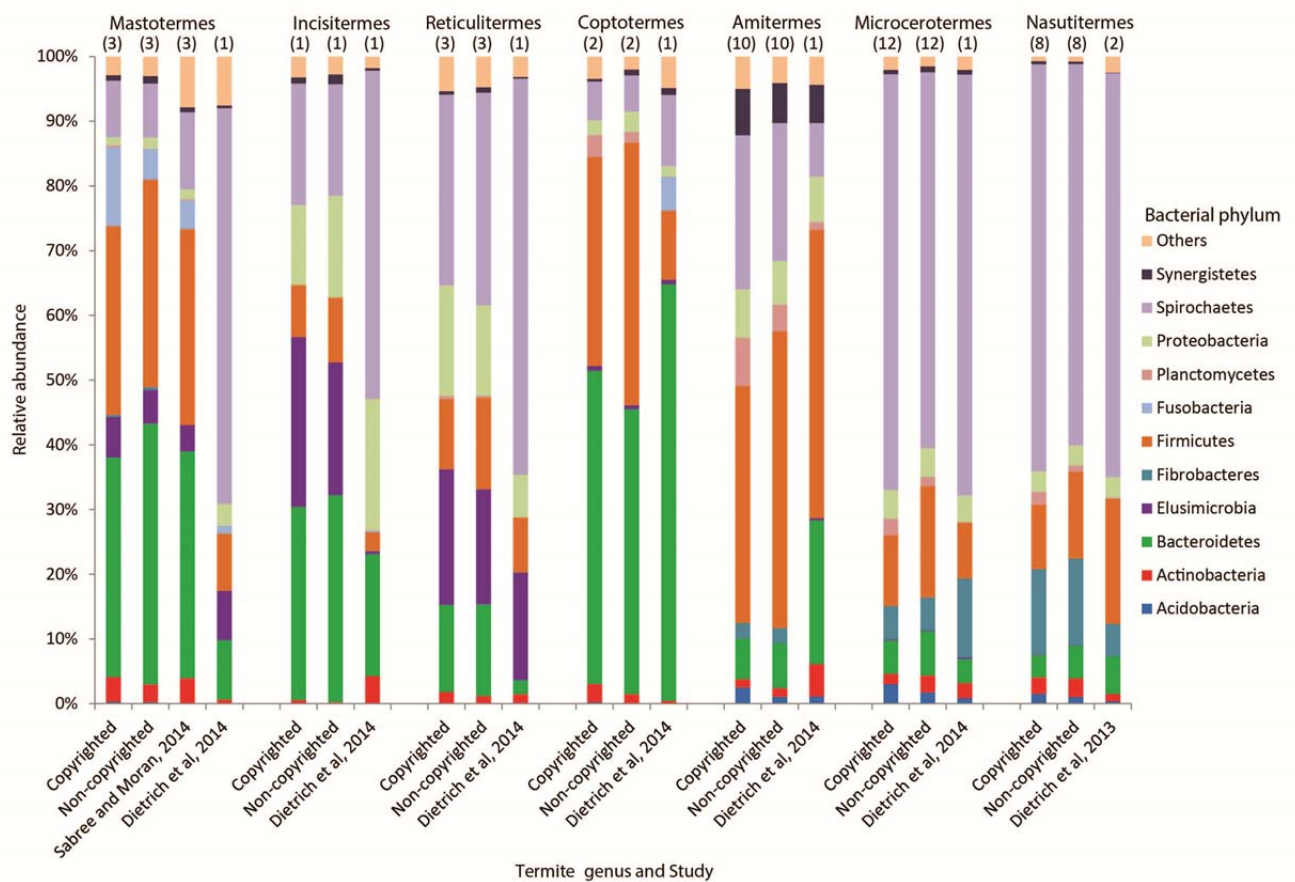


Figure S2.12: Comparison of termite gut bacterial profiles obtained in the present study (rRNA copy number corrected and uncorrected profiles) and by Dietrich et al. (2014). An additional *Mastotermes* profile reported by Sabree and Moran (2014) is also included for reference. For each study, the profiles are averaged across samples belonging to the same termite genus (number of samples is shown above each bar).

Table S2.1: Details of the 68 specimens collected in the present study.

No	ID	Collection Date	Location	GPS Coordinates	Nest Type	Presumptive Diet	$\delta^{13}C$ (%)	Family	Genus	#of workers used for profiling	Accession no
1	DR01	6/08/2011	Darwin River	S12° 46' 17.22" E130° 57' 53.58"	Free standing mound	Grass		Termitidae	Drepanotermes	30	KJ907786
2	FC04	6/08/2011	Fly Creek, Darwin	S12° 46' 19.5" E130° 58' 35.04"	Magnetic mound	Grass	-19.5	Termitidae	Amitermes	30	KJ907787
3	TV01	26/05/2011	Townsville	S19° 39' 59.18" E146° 50' 16.20"	Mound	Grass	-15.1	Termitidae	Amitermes	30	KJ907788
4	2AM2	23/07/2008	Arizona	N31° 47' 58.98" W110° 9' 26.039"	Dung	Dung		Termitidae	Gnathamitermes	30	KJ907789
5	4GP2	23/07/2008	Arizona	N31° 31' 58.74" W110° 0' 42.599"	Dung	Dung	-22.4	Termitidae	Gnathamitermes	30	KJ907790
6	7GP3	24/07/2008	Arizona	N31° 20' 52.62" W110° 14' 41.459"	Under rock	Grass	-21.6	Termitidae	Gnathamitermes	30	KJ907791
7	7GP2	24/07/2008	Arizona	N31° 20' 52.62" W110° 14' 41.459"	Under rock	Grass	-21.5	Termitidae	Gnathamitermes	30	KJ907792
8	2GP1	23/07/2008	Arizona	N31° 31' 58.74" W110° 0' 42.599"	Dung	Dung	-19.4	Termitidae	Gnathamitermes	30	KJ907793
9	7GP1	24/07/2008	Arizona	N31° 20' 52.62" W110° 14' 41.459"	Under rock	Grass	-20.7	Termitidae	Gnathamitermes	30	KJ907794
10	10GP1	24/07/2008	Arizona	N31° 20' 52.62" W110° 14' 41.459"	Under rock	Unknown		Termitidae	Gnathamitermes	30	KJ907795
11	6GP1	24/07/2008	Arizona	N31° 31' 58.74" W110° 0' 42.599"	Cow Dung on dirt	Dung	-16.9	Termitidae	Gnathamitermes	30	KJ907796
12	2AW1	23/07/2008	Arizona	N31° 47' 58.98" W110° 9' 26.039"	Wood debri	Wood	-23.2	Termitidae	Amitermes	30	KJ907797
13	4AW3	23/07/2008	Arizona	N31° 31' 58.74" W110° 0' 42.599"	Dead Yucca	Grass	-18.7	Termitidae	Amitermes	30	KJ907798
14	8AW2	23/07/2008	Arizona	N31° 31' 58.74" W110° 0' 42.599"	Cow Dung	Dung	-20.4	Termitidae	Amitermes	30	KJ907799
15	6GP2	24/07/2008	Arizona	N31° 31' 58.74" W110° 0' 42.599"	Dry wood	Wood	-23.4	Termitidae	Amitermes	30	KJ907800
16	4AW1	23/07/2008	Arizona	N31° 31' 58.74" W110° 0' 42.599"	Cow Dung	Dung	-22.4	Termitidae	Amitermes	30	KJ907801
17	4AW2	23/07/2008	Arizona	N31° 31' 58.74" W110° 0' 42.599"	Cow Dung	Dung	-20.3	Termitidae	Amitermes	30	KJ907802

Table S2.1: Continued

No	ID	Collection Date	Location	GPS Coordinates	Nest Type	Presumptive Diet	$\delta^{13}C$ (‰)	Family	Genus	#of workers used for profiling	Accession no
18	2A.1	23/07/2008	Arizona	N31° 47' 58.98" W110° 9' 26.039"	Dead root crown	Wood	-22.2	Termitidae	Amitermes	30	KJ907803
19	2A.2	23/07/2008	Arizona	N31° 47' 58.98" W110° 9' 26.039"	Soil under rock	Wood	-22.4	Termitidae	Amitermes	30	KJ907804
20	FC05	6/08/2011	Fly Creek, Darwin	S12° 46' 16.98" E130° 58' 33.24"	Cathedral mound	Grass	-13.2	Termitidae	Nasutitermes	30	KJ907805
21	DW01	18/06/2011	Darwin	S12° 28' 59.99" E130° 58' 59.99"	Cathedral mound	Grass	-14	Termitidae	Nasutitermes	30	KJ907806
22	DW04	7/08/2011	Darwin Beach	S12° 27' 46.128" E130° 50' 30.3684"	Beach Hibiscus	Wood	-28	Termitidae	Nasutitermes	30	KJ907807
23	MC05	21/07/2011	Murphy's Creek	S27° 28' 7.08" E152°01' 40.86"	Mound	Grass	-16.8	Termitidae	Nasutitermes	30	KJ907808
24	MC07	21/07/2011	Murphy's Creek	S27° 28' 12.96" E152°01' 56.64"	Mound	Grass	-18.6	Termitidae	Nasutitermes	30	KJ907809
25	MC06	21/07/2011	Murphy's Creek	S27° 28' 10.14" E152°01' 43.92"	Mound	Grass	-17.5	Termitidae	Nasutitermes	30	KJ907810
26	CC02	26/04/2011	Caloundra Coast	S26° 48' 3.6" E153° 08' 59.04"	Banksia	Wood	-27.1	Termitidae	Nasutitermes	30	KJ907811
27	LabNasut	-	Florida	N29° 38' 42.95" W82° 20' 47.018"	Laboratory	Wood	-27.4	Termitidae	Nasutitermes	30	KJ907812
28	7TT2	24/07/2008	Arizona	N31° 20' 52.62" W110° 14' 41.459"	Under rock	Grass		Termitidae	Tenuirostritermes	30	KJ907813
29	7TT3	24/07/2008	Arizona	N31° 20' 52.62" W110° 14' 41.459"	Under rock	Grass		Termitidae	Tenuirostritermes	30	KJ907814
30	CA01	10/07/2011	Cairns	S16° 52' 45" E145°44' 35"	Dead log	Wood		Termitidae	Microcerotermes	30	KJ907815
31	CA03	10/07/2011	Cairns	S16° 49' 50" E142°40' 00"	Mound	Wood		Termitidae	Microcerotermes	30	KJ907816
32	IN01	1/07/2011	Indooroopilly, Brisbane	S27° 29' 52.1916" E152° 58' 21.1944"	Nest	Wood		Termitidae	Microcerotermes	30	KJ907817
33	PH01	3/03/2011	Pinjarra Hills, QLD	S27° 32' 17.4012" E152° 55' 10.02"	Nest on post	Wood	-25.9	Termitidae	Microcerotermes	30	KJ907818
34	PH03	3/03/2011	Pinjarra Hills	S27° 32' 18.6" E152° 55' 9"	Eucalypt tree	Wood		Termitidae	Microcerotermes	30	KJ907819
35	PH02	3/03/2011	Pinjarra Hills	S27° 32' 19.2" E152° 55' 10.2"	Mound at base of Eucalypt tree	Wood		Termitidae	Microcerotermes	30	KJ907820
36	GHR03	8/06/2011	Green Hill Reservoir	S27° 29.814' E152° 57.532'	Stump	Wood		Termitidae	Microcerotermes	30	KJ907821
37	GHR01	8/06/2011	Green Hill Reservoir	S27° 30.064' E152°57.946'	Mound	Wood		Termitidae	Microcerotermes	30	KJ907822
38	MC08	21/07/2011	Murphy's Creek	S27° 28' 10.02" E152°01' 42.96"	Mound	Wood		Termitidae	Microcerotermes	30	KJ907823

Table S2.1: Continued

No	ID	Collection Date	Location	GPS Coordinates	Nest Type	Presumptive Diet	$\delta^{13}C$ (‰)	Family	Genus	#of workers used for profiling	Accession no
39	MC09	21/07/2011	Murphy's Creek	S27° 28' 14.34" E152°01' 43.5"	Mound	Wood		Termitidae	Microcerotermes	30	KJ907824
40	MC04	21/07/2011	Murphy's Creek	S27° 28' 11.34" E152°01' 41.34"	Mound	Wood		Termitidae	Microcerotermes	30	KJ907825
41	MC03	21/07/2011	Murphy's Creek	S27° 28' 6.48" E152°01' 40.14"	Mound	Wood		Termitidae	Microcerotermes	30	KJ907826
42	FC02	6/08/2011	Fly Creek, Darwin	S12° 45' 58.08" E131° 04' 55.86"	Mound at base of red gum	Wood		Termitidae	Macrognathotermes	30	KJ907827
43	4RT1	23/07/2008	Arizona	N31° 31' 58.74" W110° 0' 42.599"	Dung (probably horse)	Dung		Rhinotermitidae	Reticulitermes	30	KJ907828
44	4RT2	23/07/2008	Arizona	N31° 31' 58.74" W110° 0' 42.599"	Wood (mesquite, palo verde)	Wood		Rhinotermitidae	Reticulitermes	30	KJ907829
45	8RT1	24/07/2008	Arizona	N31° 24' 42.12" W110° 54' 23.399"	Harwood mesquite stick	Wood		Rhinotermitidae	Reticulitermes	30	KJ907830
46	MC02	21/07/2011	Murphy's Creek	S27° 28' 5.88" E152°01' 40.8"	Mound diffuse in large tree	Wood		Rhinotermitidae	Heterotermes	30	KJ907831
47	SL01	22/05/2011	St Lucia	S27° 29' 41.118" E152° 59' 37.9638"	Pine fence paling	Wood		Rhinotermitidae	Heterotermes	30	KJ907832
48	IN02	18/01/2012	Indooroopilly	S27° 30' 27.126" E152° 59' 06.252"	Roots of Oleander bush	Roots		Rhinotermitidae	Heterotermes	30	KJ907833
49	MC01	21/07/2011	Murphy's Creek	S27° 28' 6.78" E152°01' 40.98"	Tea tree root	Wood		Rhinotermitidae	Heterotermes	30	KJ907834
50	PH04	13/04/2011	Pinjarra Hills	S27° 32' 23.4" E152° 55' 8.4"	Growth of tree	Wood		Rhinotermitidae	Heterotermes	30	KJ907835
51	BF01	8/08/2011	Berrimah Farm, Darwin	S12° 26' 38.58" E130° 55' 42.9"	Pine plank with undercoat	Wood		Rhinotermitidae	Heterotermes	30	KJ907836
52	AP01	4/03/2013	Archerfield	S27° 34' 2.2116" E153 1' 1.7976"	Laboratory	Wood		Rhinotermitidae	Coptotermes	30	KJ907837
53	FC03	6/08/2011	Fly Creek, Darwin	S12° 45' 54.66" E131° 03' 4.32"	Mound at base of blood wood eucalypt	Wood		Rhinotermitidae	Coptotermes	30	KJ907838
54	BF02	8/08/2011	Berrimah Farm, Darwin	S12° 27' 25.5" E130° 55' 31.68"	Mound at base of stringy bark eucalypt	Wood		Rhinotermitidae	Coptotermes	30	KJ907839
55	CC01	26/04/2011	Caloundra Coast	S26° 48' 3.6" E153° 08' 59.04"	Banksia	Wood		Rhinotermitidae	Schedorhinotermes	20	KJ907840
56	DR02	6/08/2011	Darwin River	S12° 46' 17.22" E130° 57' 53.58"	Burnt iron wood branch	Wood		Rhinotermitidae	Schedorhinotermes	20	KJ907841
57	WH01	19/06/2011	Wellers Hill, Tarragindi	S27°31'42.21" E153° 2'53.53"	Mound	Wood		Rhinotermitidae	Schedorhinotermes	20	KJ907842
58	8MH1	24/07/2008	Arizona	N31° 35' 42" W 111° 2' 57.18"	Cottonwood	Wood		Kalotermitidae	Marginitermes	30	KJ907843
59	9MH1	24/07/2008	Arizona	N31° 47' 58.98" W110° 9' 26.039"	Cottonwood	Wood		Kalotermitidae	Marginitermes	30	KJ907844

Table S2.1: Continued

No	ID	Collection Date	Location	GPS Coordinates	Nest Type	Presumptive Diet	$\delta^{13}\text{C}$ (‰)	Family	Genus	#of workers used for profiling	Accession no
60	SIM1	24/07/2008	Arizona	N31° 31' 58.74" W110° 0' 42.599"	Sycamore	Wood		Kalotermitidae	Incisitermes	30	KJ907845
61	BB01-2	4/03/2011	Binna Burra	S28° 9' 29.916" E153° 11' 21.771"	Dead tree	Wood		Kalotermitidae	Glyptotermes	30	KJ907846
62	TN05	19/11/2011	Mount Glorious	S27° 20' 6.96" E152° 46' 13.5"	Fallen log	Wood		Kalotermitidae	Glyptotermes	30	KJ907847
63	TN01	19/11/2011	Mount Glorious	S27° 20' 11.6406" E152° 46' 13.2708"	Fallen log	Wood		Stolotermitidae	Porotermes	30	KJ907848
64	FC01	6/08/2011	Fly Creek, Darwin	S12° 45' 58.05" E131° 04' 55.86"	Cattle dung / soil	Dung	-26.4	Mastotermitidae	Mastotermes	5	KJ907849
65	DW03	5/08/2011	Darwin, Northern Suburbs	S12° 22' 55.5" E130° 53' 55.1"	Carpentaria palm	Grass	-27.0	Mastotermitidae	Mastotermes	5	KJ907850
66	DW02	5/08/2011	Darwin, Northern Suburbs	S12° 22' 55.1" E130° 53' 53.7"	African mahogany	Wood	-27.5	Mastotermitidae	Mastotermes	5	KJ907851
67	PP01	22/07/2011	Perrin Park, Toowong	S27° 29' 32.5824" E153° 0' 25.2168"	Palm tree	Omnivorous		Blattidae	Periplaneta	1	KJ907852
68	DR03	6/08/2011	Darwin River	S12° 46' 17.22" E130° 57' 53.58"	Under Bark	Omnivorous		Blattidae	Periplaneta	1	KJ907853

Table S2.2: Effect of rRNA copy number correction on termite gut profile relative abundance estimates for OTUs with $\geq 1\%$ in at least one sample.

<http://link.springer.com/article/10.1186%2Fs40168-015-0067-8>

Table S2.3: Distribution of identical sequence clusters comprising >10 reads in *Treponema* OTU1 (comprising 7,223 reads in total). Each row represents an identical cluster and each column a gut sample with absolute numbers of reads for each cluster and sample shown. The termite genus and family for each sample and country of origin are indicated at the top of the table using color coding.

<http://link.springer.com/article/10.1186%2Fs40168-015-0067-8>

Table S2.4: Consistency analysis of microbial community relationships based on weighted and unweighted Soergel distances with host phylogeny (COII) and presumptive diet.

Attribute	Unique states	# state changes		Consistency index	
		unweighted	weighted	unweighted	weighted
<i>Lower and Higher Termites</i>					
Family	5	5	9	0.8	0.44
Genus	17	17	21	0.94	0.76
Diet	3	15	12	0.13	0.17
<i>Higher Termites</i>					
Genus	7	6	7	1	0.86
Diet	3	15	12	0.13	0.17

Appendix B: Supplementary figures and tables for Chapter 3

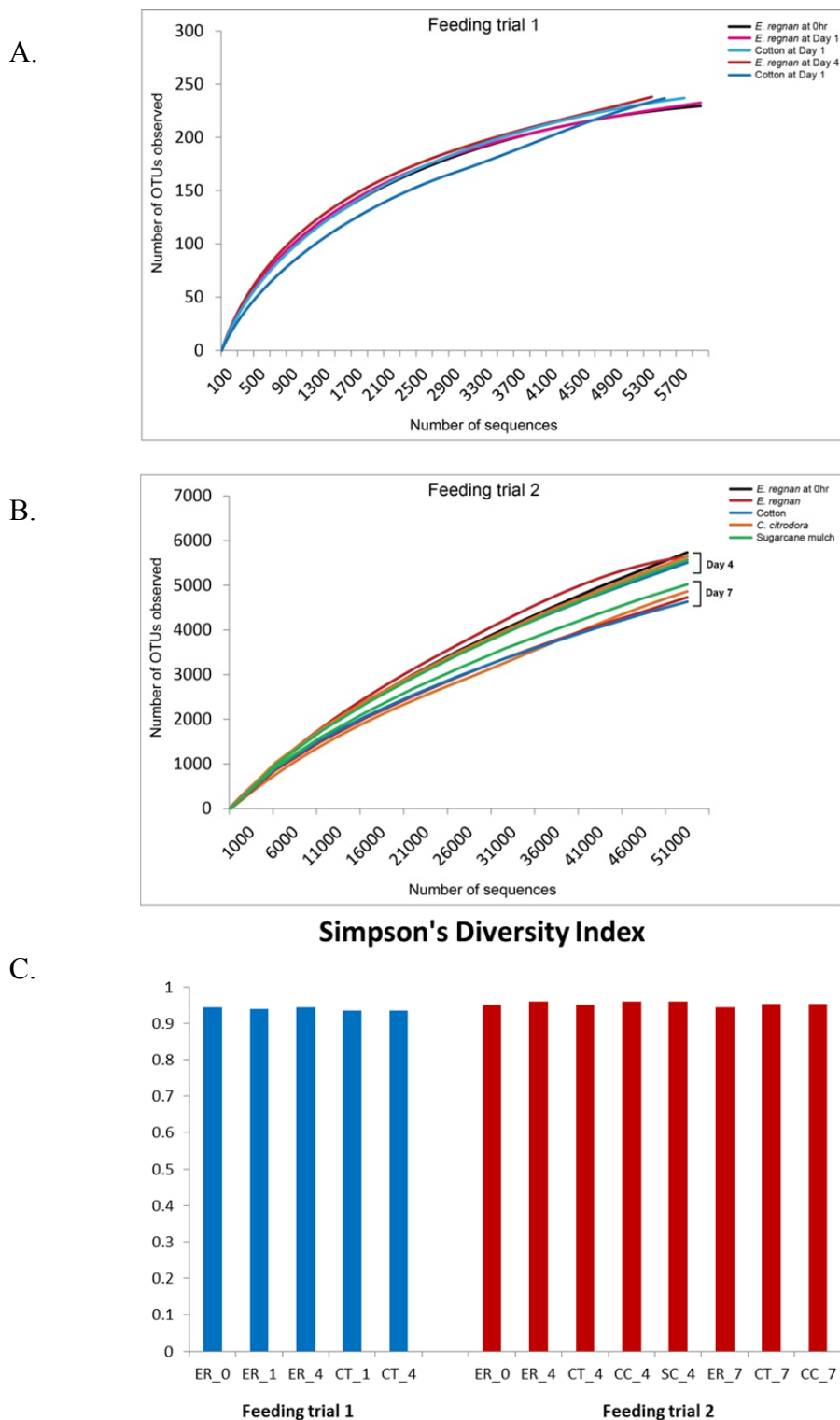


Figure S3.1: Rarefaction curves and associated Simpson diversity index of microbial profiles obtained for each biological replicates. (A) Alpha diversity analysis for feeding trial 1, (B) feeding trial 2. (C) Simpson's diversity index for feeding trial 1 (blue) and 2 (red).

Figure S3.2: Total protein concentration of the obtained crude extract from different feeding treatments.

<i>E. regnans</i> at 0hr									
<i>Average Net595</i>			0.8273	0.5337	0.3145	0.8618	0.5739	0.3019	
0.7824	0.48065	0.26955							
<i>Concentration ug/ml</i>			1620.704	707.6507	375.1442	1815.58	783.4437	357.5178	
1404.778	618.3353	312.4503							
<i>ug/ml * dilution</i>			12965.63	11322.41	12004.61	14524.64	12535.1	11440.57	
11238.23	9893.365	9998.41							
<i>mg/ml</i>			12.96563	11.32241	12.00461	14.52464	12.5351	11.44057	
11.23823	9.893365	9.99841							
<i>Concentration mg/ml</i>									
ER1			ER2			ER3			
10.37667			12.09755			12.83344			

<i>E. regnans</i> at Day 7									
<i>Average Net595</i>			0.59065	0.37115	0.18995	0.41985	0.2406	0.1138	
0.61345	0.4361	0.2344							
<i>Concentration ug/ml</i>			817.7399	455.0653	203.5079	525.474	272.4012	106.306	
867.4977	549.601	263.8721							
<i>ug/ml * dilution</i>			6541.919	7281.045	6512.254	4203.792	4358.419	3401.792	
6939.981	8793.616	8443.908							
<i>mg/ml</i>			6.541919	7.281045	6.512254	4.203792	4.358419	3.401792	
6.939981	8.793616	8.443908							
<i>Concentration mg/ml</i>									
ER1			ER2			ER3			
8.059169			6.778406			3.988001			

<i>C. Citroedora</i> at Day 7									
<i>Average Net595</i>			0.4676	0.26695	0.14815	0.6365	0.3813	0.2236	
0.57185	0.33645	0.19735							
<i>Concentration ug/ml</i>			568.0068	334.7553	137.1801	871.2417	475.8851	266.8214	
723.2108	426.2546	222.4307							
<i>ug/ml * dilution</i>			4544.055	5356.085	4389.763	6969.933	7614.161	8538.283	
5785.686	6820.074	7117.781							
<i>mg/ml</i>			4.544055	5.356085	4.389763	6.969933	7.614161	8.538283	
5.785686	6.820074	7.117781							
<i>Concentration mg/ml</i>									
CC1			CC2			CC3			
6.574514			4.763301			7.707459			

Figure S3.2: Continued.

Cotton at Day 7								
<i>Average Net595</i>								
0.46555	0.20275	0.16885	0.78995	0.4684	0.29845	0.5124	0.28265	0.1471
<i>Concentration ug/ml</i>								
516.2334	217.8478	189.1229	1311.54	521.0186	301.8475	600.3527	286.9238	169.8896
269.5918								
<i>ug/ml * dilution</i>								
4129.867	3485.564	6051.933	10492.32	8336.298	9659.119	4802.822	4590.78	5436.468
4313.469								
<i>mg/ml</i>								
4.129867	3.485564	6.051933	10.49232	8.336298	9.659119	4.802822	4.59078	5.436468
4.313469								
<i>Concentration mg/ml</i>								
CT1			CT2			CT3		
4.221668			9.495912			4.943357		

Sugarcane mulch at Day 7								
<i>Average Net595</i>								
0.889	0.5322	0.3648	0.6815	0.4034	0.231	0.547	0.3365	0.19485
<i>Concentration ug/ml</i>								
1607.702	639.4271	372.9113	997.8388	422.1455	241.5417	670.0006	340.7278	211.227
<i>ug/ml * dilution</i>								
12861.62	10230.83	11933.16	7982.711	6754.328	7729.334	5360.005	5451.645	6759.265
<i>mg/ml</i>								
12.86162	10.23083	11.93316	7.982711	6.754328	7.729334	5.360005	5.451645	6.759265
<i>Concentration mg/ml</i>								
SC1			SC2			SC3		
11.6752			7.488791			5.856972		

Table S3.1: Relative abundance in the mitag libraries of the *Mastotermes* gut microbiota.

#OTU ID	ER1_0	ER2_0	ER3_0	ER1_4	ER2_4	ER3_4	ER1_7	ER2_7	ER3_7	CC1_4	CC2_4	CC3_4	CC1_7	CC2_7	CC3_7
OTU131;Deltotrichonympha_nana	1.4	2.0	2.2	1.4	3.0	1.3	0.6	0.5	0.4	2.9	0.8	1.1	1.9	0.1	0.1
OTU77;Koruga_bonita	0.1	0.2	0.2	0.2	0.3	0.2	0.0	0.0	0.0	0.2	0.2	0.2	0.2	0.0	0.0
OTU161;Metadevescovina_extranea	0.1	0.2	0.2	0.3	0.5	0.1	0.1	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Parabasalia	1.5	2.4	2.6	1.9	3.9	1.6	0.7	1.0	0.5	3.1	1.0	1.3	2.1	0.1	0.2
OTU86;f_Methanobacteriaceae	0.3	0.2	0.3	0.2	0.1	0.2	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1
OTU28;g_Methanobacterium	0.4	0.4	0.6	0.3	0.0	0.7	0.0	0.1	0.0	0.0	0.8	0.1	0.0	0.0	0.0
OTU9;g_Methanobacterium; s_beijingense	2.7	1.8	2.5	2.4	1.3	2.7	0.9	1.3	1.3	2.5	1.5	1.3	1.1	0.6	0.9
OTU31;g_Methanobrevibacter	0.2	0.5	0.3	0.2	0.1	0.1	0.0	0.3	0.1	0.1	0.1	0.0	0.1	1.5	0.0
OTU38;g_Methanobrevibacter; s_arboriphilus_1	5.5	6.4	7.7	5.2	1.1	7.2	1.5	2.0	2.2	1.0	6.5	1.8	1.7	2.1	2.0
OTU89;g_Methanobrevibacter; s_arboriphilus_2	0.3	0.2	0.3	0.3	0.2	0.4	0.6	0.6	1.0	0.2	0.5	0.3	0.6	0.8	0.8
Euryarchaeota	9.4	9.4	11.7	8.5	2.8	11.4	3.2	4.3	4.6	4.0	9.4	3.7	3.6	5.1	3.9
OTU113;p_Acidobacteria; o_MVS-40	0.5	0.5	0.6	0.7	0.6	0.7	0.8	0.6	0.5	0.7	0.4	0.6	0.6	0.5	0.3
Acidobacteria	0.5	0.5	0.6	0.7	0.6	0.7	0.8	0.6	0.5	0.7	0.4	0.6	0.6	0.5	0.3
OTU52;g_Corynebacterium	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.1	0.1	5.4	0.3
OTU51;g_Microbacterium_2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1
OTU55;g_Microbacterium_3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
OTU147;g_Microbacterium_1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.4	0.0	0.0	0.0	0.0	0.2	0.0
OTU10;g_Propionicimonas_2	0.3	0.1	0.1	0.3	0.1	0.2	0.2	0.2	0.3	0.1	0.1	0.2	0.2	0.2	0.3
OTU12;g_Propionicimonas_3	0.1	0.1	0.1	0.3	0.1	0.3	0.1	0.2	0.3	0.1	0.2	0.2	0.2	0.3	0.1
OTU53;g_Propionicimonas_1	0.2	0.1	0.1	0.2	0.4	0.1	0.1	0.2	0.3	0.1	0.1	0.1	0.2	0.1	0.3
OTU73;g_Xylanimicrobium	0.1	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
OTU24;f_Propionibacteriaceae	1.0	0.5	1.1	1.3	0.9	1.5	3.3	2.4	3.9	1.1	0.8	1.5	2.3	3.4	4.4
OTU42;g_Tsukamurella	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
OTU23;f_Coriobacteriaceae_1	0.2	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.1	0.1	0.2	0.1	0.3
OTU37;f_Coriobacteriaceae_2	0.3	0.2	0.3	0.5	0.2	0.3	0.2	0.4	0.3	0.2	0.3	0.2	0.3	0.4	0.4
Actinobacteria	1.9	1.2	1.9	2.7	1.9	2.6	4.4	4.0	6.2	1.9	1.7	2.4	3.8	10.5	6.3
OTU134;o_Bacteroidales_1	0.4	0.2	0.2	0.4	0.2	0.3	0.1	0.6	0.2	0.2	0.3	0.2	0.4	0.2	0.1
OTU17;o_Bacteroidales_4	0.1	0.2	0.1	0.2	0.1	0.4	0.1	0.4	0.8	0.2	0.3	0.2	0.5	0.3	0.3
OTU72;o_Bacteroidales_6	0.6	1.2	0.7	1.7	0.3	1.4	0.1	0.3	0.4	1.2	1.2	1.5	0.6	0.2	0.3
OTU82;o_Bacteroidales_3	0.3	0.6	0.4	0.7	0.1	0.6	0.1	0.2	0.2	0.6	0.6	0.8	0.3	0.1	0.2
OTU114;o_Bacteroidales_7	0.2	0.2	0.3	0.1	0.2	0.3	0.1	0.0	0.4	0.3	0.3	0.3	0.3	0.2	0.3
OTU159;o_Bacteroidales_2	0.3	0.2	0.3	0.2	0.2	0.3	0.1	0.1	0.1	0.3	0.3	0.4	0.2	0.0	0.1
OTU160;o_Bacteroidales_5	0.0	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1
OTU95;f_Bacteroidaceae	0.3	0.4	0.3	0.6	0.2	0.4	0.1	0.3	0.3	0.3	0.3	0.2	0.3	0.1	0.1
OTU22;g_Bacteroides_2	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.2	0.2	0.0	0.0	0.0	0.1	0.1	0.0
OTU63;g_Bacteroides_3	0.1	0.0	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.1
OTU145;g_Bacteroides_1	0.1	0.1	0.2	0.2	0.3	0.1	0.0	0.3	0.2	0.1	0.1	0.1	0.1	0.3	0.0
OTU20;g_Bacteroides; s_eggerthii	0.4	0.1	0.1	0.2	1.3	0.4	0.1	0.5	0.2	0.2	0.1	0.2	0.5	0.7	0.2
OTU142;g_Bacteroides; s_uniformis	0.2	0.1	0.2	0.1	0.1	0.3	0.1	0.1	0.1	0.2	0.2	0.3	0.2	0.1	0.2
OTU120;o_Bacteroidales; f_p-2534-18B5_2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.4	0.2	0.5	0.1	0.0	0.1
OTU137;o_Bacteroidales; f_p-2534-18B5_1	1.5	1.0	1.5	0.7	0.8	1.2	0.8	0.3	0.6	3.9	2.0	3.6	0.9	0.3	0.9
OTU49;f_Porphryomonadaceae	0.2	0.1	0.1	0.2	0.4	0.1	0.3	0.3	0.3	0.2	0.1	0.2	0.4	0.3	0.5
OTU19;g_Candidatus Azobacteroides_4	3.3	2.7	3.2	3.3	1.6	2.8	2.5	2.1	3.3	3.0	2.8	3.4	2.5	2.3	2.3
OTU41;g_Candidatus Azobacteroides_2	0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.1	0.1	0.2	0.2	0.2	0.2	0.1	0.2
OTU111;g_Candidatus Azobacteroides_1	0.2	0.2	0.1	0.1	0.1	0.2	0.1	0.1	0.2	0.1	0.1	0.1	0.2	0.1	0.1
OTU124;g_Candidatus Azobacteroides_5	0.6	0.5	0.6	0.6	0.3	0.5	0.5	0.4	0.6	0.6	0.6	0.7	0.5	0.6	0.5

Table S3.1: Continued.

#OTU ID	ER1_0	ER2_0	ER3_0	ER1_4	ER2_4	ER3_4	ER1_7	ER2_7	ER3_7	CC1_4	CC2_4	CC3_4	CC1_7	CC2_7	CC3_7
OTU128;g_Candidatus Azobacteroides_3	1.2	2.2	1.3	1.2	1.3	2.1	1.1	1.0	1.6	2.1	1.2	2.0	2.0	1.0	2.5
OTU74;g_Dysgonomonas_4	0.0	0.2	0.1	0.1	0.1	0.0	0.0	1.3	0.1	0.1	0.0	0.0	0.0	0.2	0.0
OTU78;g_Dysgonomonas_3	0.5	0.4	0.4	0.5	0.3	0.5	0.1	0.1	0.1	0.3	0.4	0.3	0.3	0.1	0.1
OTU148;g_Dysgonomonas_1	2.6	3.8	3.9	8.0	4.3	3.4	5.5	11.2	9.9	6.8	4.6	5.3	9.0	7.1	3.8
OTU158;g_Dysgonomonas_2	0.1	0.1	0.1	0.3	0.1	0.2	0.2	0.4	0.2	0.4	0.2	0.3	0.4	0.2	0.3
OTU40;g_Parabacteroides	0.3	0.2	0.3	0.3	0.5	0.3	0.4	0.3	0.4	0.3	0.3	0.4	0.3	0.3	0.6
OTU59;g_Parabacteroides	0.3	0.2	0.3	0.2	0.2	0.3	0.4	0.2	0.2	0.3	0.3	0.3	0.3	0.5	0.5
OTU68;g_Parabacteroides	2.3	1.5	2.2	2.3	2.0	2.5	3.4	2.3	1.9	2.7	2.5	2.9	2.6	2.9	3.8
OTU71;g_Parabacteroides	2.1	1.9	2.1	2.3	1.2	2.2	1.8	1.2	1.7	2.0	2.2	2.5	1.7	1.3	1.8
OTU85;g_Parabacteroides	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
OTU88;g_Parabacteroides	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.1
OTU98;g_Parabacteroides	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1
OTU99;g_Parabacteroides	0.1	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0
OTU115;g_Parabacteroides	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.2	0.2	0.1	0.2	0.2
OTU126;g_Parabacteroides	0.5	0.5	0.6	0.5	0.3	0.6	0.5	0.3	0.5	0.6	0.6	0.6	0.5	0.5	0.5
OTU139;g_Parabacteroides	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.0
OTU67;g_Tannerella	2.9	1.9	2.7	2.3	2.4	3.1	3.0	2.1	1.8	2.6	2.1	3.0	2.6	2.8	2.8
OTU75;g_Tannerella	0.8	0.9	0.9	0.5	0.1	0.7	0.0	0.1	0.0	0.1	0.7	0.1	0.0	0.0	0.0
OTU143;g_Tannerella	0.6	0.4	0.5	0.3	0.2	0.8	0.4	0.2	0.1	0.5	0.4	0.6	0.4	0.0	0.1
OTU125;g_Prevotella	0.3	0.2	0.3	0.1	0.3	0.3	0.3	0.2	0.2	1.0	0.4	0.4	0.3	0.3	0.4
OTU32;o_Bacteroidales; f_RF16_1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.2	0.1	0.1	0.1
OTU90;o_Bacteroidales; f_RF16_2	0.1	0.1	0.1	0.2	0.1	0.1	0.0	0.1	0.1	0.2	0.1	0.2	0.1	0.0	0.0
OTU153;o_Bacteroidales; f_RF16_3	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.4	0.0	0.1	0.0	0.1	0.1	0.0
OTU132;o_Bacteroidales; f_S24-7	0.1	0.1	0.2	0.3	0.9	0.2	0.6	1.6	2.5	0.2	0.3	0.3	0.7	1.2	0.6
OTU61;g_Chryseobacterium	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
OTU57;g_Blattabacterium	12.9	14.3	8.9	5.6	14.1	6.4	15.4	18.2	14.4	2.9	14.5	4.3	4.9	4.3	17.7
OTU84;g_Blattabacterium	0.3	0.3	0.2	0.2	0.3	0.2	0.3	0.4	0.3	0.1	0.3	0.2	0.1	0.1	0.3
OTU140;g_Blattabacterium	0.2	0.2	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.0	0.1	0.1	0.1	0.1	0.4
OTU151;g_Blattabacterium	0.1	0.2	0.1	0.1	0.2	0.1	0.2	0.2	0.2	0.0	0.2	0.1	0.1	0.1	0.2
OTU155;g_Blattabacterium	0.3	0.4	0.2	0.2	0.3	0.2	0.4	0.4	0.3	0.1	0.3	0.2	0.1	0.2	0.4
OTU60;g_Sphingobacterium; s_multivorum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0
Bacteroidetes	38.2	38.5	35.0	35.9	36.0	34.4	40.6	49.1	46.3	36.1	42.3	37.9	35.6	30.1	44.1
OTU26;c_Endomicrobia_5	0.5	0.5	0.3	0.3	0.5	0.3	0.2	0.2	0.1	0.3	0.3	0.4	0.2	0.0	0.1
OTU27;c_Endomicrobia_3	14.9	12.8	11.3	7.7	14.3	9.3	2.5	2.1	1.3	15.4	3.5	9.1	6.1	0.1	1.1
OTU81;c_Endomicrobia_2	0.3	0.2	0.2	0.2	0.3	0.2	0.0	0.0	0.0	0.3	0.1	0.2	0.1	0.0	0.0
OTU92;c_Endomicrobia_4	0.6	0.5	0.4	0.4	0.8	0.3	0.0	0.1	0.0	0.3	0.2	0.3	0.2	0.0	0.0
OTU102;c_Endomicrobia_7	0.7	0.9	0.4	0.2	0.7	0.1	0.3	0.2	0.1	1.3	0.2	0.8	0.5	0.1	0.2
OTU105;c_Endomicrobia_8	0.3	0.2	0.2	0.2	0.3	0.2	0.1	0.1	0.0	0.3	0.1	0.2	0.1	0.0	0.0
OTU122;c_Endomicrobia_1	1.5	1.3	1.6	1.1	0.1	1.6	0.1	0.1	0.0	0.1	1.3	0.2	0.0	0.0	0.0
OTU144;c_Endomicrobia_6	1.5	0.9	2.0	0.7	0.5	1.7	0.8	0.4	0.1	1.8	1.2	2.4	0.5	0.0	0.1
Elusimicrobia	20.2	17.4	16.4	11.0	17.3	13.8	4.1	3.1	1.7	19.7	6.9	13.6	7.7	0.3	1.6
OTU36;g_Enterococcus	0.6	1.3	1.0	1.5	0.6	2.4	0.4	0.3	0.3	1.1	1.0	0.8	0.3	0.8	0.7
OTU29;f_Leuconostocaceae_1	0.9	2.4	1.9	4.2	1.9	2.5	1.8	0.9	1.0	2.7	2.1	2.2	2.3	1.4	1.4
OTU70;f_Leuconostocaceae_2	4.2	7.3	5.6	8.2	8.8	6.6	5.5	8.1	10.7	9.0	10.7	8.1	9.0	14.7	8.8
OTU100;f_Leuconostocaceae_3	0.5	1.4	1.2	2.7	1.0	1.5	1.2	0.7	0.7	1.5	1.1	1.4	1.4	0.7	0.8
OTU14;g_Lactococcus_4	0.3	0.4	0.4	0.5	1.1	0.3	0.7	0.4	0.5	0.3	0.6	0.1	0.1	0.7	0.2

Table S3.1: Continued.

#OTU ID	ER1_0	ER2_0	ER3_0	ER1_4	ER2_4	ER3_4	ER1_7	ER2_7	ER3_7	CC1_4	CC2_4	CC3_4	CC1_7	CC2_7	CC3_7
OTU35;g_Lactococcus_2	0.1	0.0	0.1	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
OTU50;g_Lactococcus_3	0.0	0.1	0.1	0.2	0.0	0.2	0.1	0.0	0.1	0.1	0.2	0.1	0.1	0.0	0.1
OTU108;g_Lactococcus_1	0.1	0.1	0.2	0.2	0.0	0.1	0.0	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.1
OTU80;o_Clostridiales_4	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.2	0.1	0.1
OTU106;o_Clostridiales_3	0.3	0.4	0.4	0.6	0.3	0.6	0.2	1.2	0.4	0.4	0.3	0.3	0.6	0.7	0.1
OTU123;o_Clostridiales_1	0.1	0.2	0.2	0.0	0.2	0.0	0.1	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
OTU138;o_Clostridiales_2	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.2	0.0	0.1	0.0	0.1	0.1	0.0	0.0
OTU141;o_Clostridiales_5	0.2	0.2	0.2	0.2	0.1	0.4	0.1	0.0	0.0	0.6	0.2	0.2	0.1	0.0	0.0
OTU18;f_Mogibacteriaceae_7	0.2	0.2	0.2	0.4	0.1	0.3	0.1	0.4	0.3	0.2	0.3	0.2	0.2	0.3	0.1
OTU25;f_Mogibacteriaceae_3	1.8	1.2	1.5	2.1	1.0	1.8	1.1	2.2	2.1	1.3	1.6	1.4	1.8	1.3	1.7
OTU45;f_Mogibacteriaceae_2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
OTU47;f_Mogibacteriaceae_9	0.2	0.3	0.3	0.5	0.4	0.1	0.1	0.2	0.2	0.2	0.1	0.1	0.1	0.2	0.1
OTU94;f_Mogibacteriaceae_5	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1
OTU101;f_Mogibacteriaceae_8	0.2	0.1	0.1	0.2	0.1	0.2	0.1	0.2	0.2	0.1	0.2	0.2	0.2	0.1	0.2
OTU129;f_Mogibacteriaceae_4	0.6	0.3	0.5	0.4	0.2	0.3	0.4	0.5	0.5	0.3	0.2	0.3	0.4	0.2	0.3
OTU150;f_Mogibacteriaceae_1	0.4	0.3	0.4	0.3	0.4	0.3	0.4	0.4	0.2	0.3	0.3	0.4	0.3	0.2	0.3
OTU154;f_Mogibacteriaceae_6	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1
OTU112;f_Clostridiaceae_2	0.1	0.1	0.2	0.5	0.1	0.8	0.8	0.2	1.0	0.3	0.3	0.6	0.2	0.7	0.7
OTU118;f_Clostridiaceae_1	0.0	0.1	0.0	0.1	0.1	0.1	0.2	0.1	0.2	0.1	0.2	0.3	0.3	0.1	0.3
OTU6;g_Clostridium; s_subterminale	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.2
OTU109;g_Anaerofustis_1	0.1	0.2	0.2	0.1	0.1	0.2	0.2	0.3	0.1	0.2	0.1	0.1	0.1	0.0	0.0
OTU110;g_Anaerofustis_2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0
OTU8;f_Lachnospiraceae_1	0.0	0.0	0.0	0.0	0.2	0.0	0.3	0.2	0.1	0.0	0.1	0.1	0.2	0.3	0.2
OTU34;f_Lachnospiraceae_2	0.0	0.0	0.0	0.1	0.2	0.1	0.2	0.8	0.2	0.0	0.2	0.2	0.7	0.5	0.6
OTU69;f_Lachnospiraceae_3	0.2	0.5	0.3	1.4	5.7	0.6	15.2	9.1	5.0	0.2	1.5	2.5	10.3	14.4	5.5
OTU56;g_Anaerostipes	0.1	0.1	0.1	0.2	0.1	0.1	0.2	0.2	0.2	0.1	0.2	0.1	0.2	0.1	0.1
OTU11;f_Peptostreptococcaceae_1	0.0	0.0	0.0	0.0	0.1	0.1	0.2	0.1	0.2	0.1	0.2	0.3	0.3	0.1	0.3
OTU48;f_Peptostreptococcaceae_5	0.1	0.1	0.1	0.1	0.1	0.2	0.4	0.2	0.3	0.2	0.4	0.5	0.7	0.3	0.6
OTU116;f_Peptostreptococcaceae_4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
OTU121;f_Peptostreptococcaceae_2	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.1
OTU136;f_Peptostreptococcaceae_3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1
OTU54;f_Ruminococcaceae_7	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.2	0.3	0.1	0.1	0.1	0.1	0.1	0.0
OTU93;f_Ruminococcaceae_2	0.2	0.2	0.3	0.3	0.1	0.3	0.2	0.1	0.2	0.1	0.2	0.1	0.1	0.1	0.1
OTU103;f_Ruminococcaceae_3	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
OTU107;f_Ruminococcaceae_8	0.8	0.6	0.9	0.9	0.2	0.6	0.1	0.2	0.1	0.2	0.2	0.1	0.2	0.0	0.0
OTU130;f_Ruminococcaceae_1	0.1	0.1	0.1	0.2	0.1	0.2	0.0	0.2	0.1	0.3	0.2	0.1	0.1	0.0	0.0
OTU133;f_Ruminococcaceae_5	0.5	0.2	0.3	0.1	0.3	0.3	0.1	0.1	0.0	0.2	0.2	0.2	0.1	0.0	0.1
OTU149;f_Ruminococcaceae_4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.2
OTU156;f_Ruminococcaceae_5	0.5	0.3	0.4	0.3	0.3	0.2	0.0	0.3	0.0	0.2	0.1	0.1	0.1	0.0	0.0
OTU46;g_Phascolarctobacterium	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.2	0.0
OTU21;f_Erysipelotrichaceae	0.0	0.2	0.1	0.4	1.0	0.1	0.3	0.2	1.7	0.1	0.1	0.3	0.6	0.8	0.8
Firmicutes	14.8	19.7	18.5	28.0	26.2	22.6	31.6	29.7	28.0	21.4	24.4	23.0	32.5	40.2	25.5
OTU96;g_Fusobacterium_2	0.1	0.1	0.1	0.2	0.2	0.2	0.3	0.1	0.2	0.2	0.2	0.2	0.3	0.4	0.3
OTU97;g_Fusobacterium_1	3.6	3.3	3.5	3.5	4.2	3.0	8.0	3.0	5.9	4.4	4.1	5.9	5.3	7.0	8.3
OTU119;g_Fusobacterium_3	0.2	0.1	0.2	0.2	0.1	0.2	0.3	0.1	0.2	0.2	0.2	0.3	0.2	0.2	0.2
Fusobacteria	3.9	3.5	3.8	3.9	4.5	3.4	8.6	3.2	6.4	4.9	4.5	6.5	5.8	7.6	8.8

Table S3.1: Continued.

#OTU ID	ER1_0	ER2_0	ER3_0	ER1_4	ER2_4	ER3_4	ER1_7	ER2_7	ER3_7	CC1_4	CC2_4	CC3_4	CC1_7	CC2_7	CC3_7
OTU66;f_Pirellulaceae	0.2	0.2	0.2	0.2	0.1	0.2	0.3	0.3	0.3	0.2	0.2	0.2	0.3	0.3	0.3
Planctomycetes	0.2	0.2	0.2	0.2	0.1	0.2	0.3	0.3	0.3	0.2	0.2	0.2	0.3	0.3	0.3
OTU79;o_Rickettsiales	1.0	0.2	1.8	0.8	0.4	0.2	0.0	0.0	0.0	0.2	0.0	0.1	0.0	0.0	0.0
OTU43;c_Betaproteobacteria	0.3	0.3	0.4	0.5	0.3	0.5	0.5	0.3	0.5	0.6	0.4	0.4	0.5	0.6	0.3
OTU65;f_Oxalobacteraceae	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.2	0.2	0.1
OTU13;f_Rhodocyclaceae; g_TS34_2	1.0	1.3	1.2	0.9	1.1	1.8	1.9	0.9	1.6	1.0	1.9	1.6	1.9	1.6	1.5
OTU135;f_Rhodocyclaceae; g_TS34_1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
OTU39;f_Enterobacteriaceae_1	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0
OTU64;f_Enterobacteriaceae_2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0
OTU76;f_Enterobacteriaceae_3	0.8	0.0	0.0	0.1	0.7	0.7	0.0	0.6	0.2	0.0	0.1	0.1	0.9	0.0	0.1
OTU44;g_Acinetobacter	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0
OTU33;g_Pseudomonas	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	3.6
OTU4;g_Stenotrophomonas; s_geniculata	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2
Proteobacteria	3.4	2.1	3.7	2.6	2.7	3.4	2.7	2.1	2.7	2.0	2.7	2.4	3.8	3.5	5.9
OTU1;g_Treponema_7	0.5	0.4	0.3	0.2	0.5	0.3	0.1	0.2	0.1	0.5	0.1	0.2	0.2	0.0	0.1
OTU2;g_Treponema_6	2.0	1.4	1.8	1.3	0.2	1.3	0.4	0.1	0.1	1.1	1.0	0.9	0.3	0.0	0.0
OTU5;g_Treponema_1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.1	0.2	0.1	0.1
OTU7;g_Treponema_10	0.1	0.2	0.1	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.2	0.1	0.1
OTU15;g_Treponema_11	0.2	0.2	0.2	0.1	0.1	0.2	0.1	0.1	0.0	0.2	0.1	0.2	0.2	0.0	0.0
OTU58;g_Treponema_9	0.3	0.2	0.3	0.3	0.2	0.4	0.0	0.0	0.0	0.5	0.1	0.3	0.0	0.0	0.0
OTU62;g_Treponema_2	0.2	0.1	0.2	0.2	0.2	0.5	0.1	0.1	0.0	0.3	0.1	0.1	0.1	0.0	0.1
OTU91;g_Treponema_3	0.2	0.2	0.1	0.1	0.2	0.2	0.2	0.1	0.1	0.2	0.2	0.2	0.2	0.0	0.1
OTU104;g_Treponema_8	0.3	0.4	0.3	0.2	0.3	0.4	0.1	0.2	0.1	0.5	0.3	0.3	0.3	0.0	0.1
OTU146;g_Treponema_4	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.1	0.2
OTU157;g_Treponema_5	0.1	0.2	0.1	0.1	0.1	0.1	0.0	0.1	0.0	0.2	0.1	0.1	0.1	0.0	0.0
OTU152;g_Treponema; s_primitia	0.2	0.2	0.2	0.1	0.3	0.2	0.1	0.1	0.1	0.2	0.1	0.2	0.1	0.0	0.1
OTU3;g_Treponema; s_sp5	0.3	0.1	0.3	0.2	0.1	0.3	0.1	0.1	0.0	0.2	0.1	0.2	0.1	0.0	0.0
Spirochaetes	4.8	4.1	4.3	3.4	2.9	4.4	1.7	1.5	1.1	4.3	2.8	3.1	2.4	0.6	1.0
OTU16;f_Dethiosulfovibrionaceae; g_TG5_2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1
OTU87;f_Dethiosulfovibrionaceae; g_TG5_1	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1
OTU30;g_Candidatus Tammella	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.2	0.1	0.1	0.1
OTU127;g_Candidatus Tammella	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.1	0.2	0.1	0.1	0.1	0.1
Synergistetes	0.3	0.3	0.3	0.3	0.3	0.4	0.3	0.1	0.3	0.3	0.4	0.4	0.4	0.3	0.3
OTU117;c_Mollicutes; o_RsaHF231	0.3	0.3	0.6	0.7	0.4	1.1	0.9	0.5	1.6	1.0	3.2	4.6	1.4	1.1	1.9
Tenericutes	0.3	0.3	0.6	0.7	0.4	1.1	0.9	0.5	1.6	1.0	3.2	4.6	1.4	1.1	1.9
OTU83;f_Cerasicoccaceae	0.6	0.4	0.5	0.3	0.3	0.2	0.1	0.2	0.0	0.4	0.2	0.2	0.1	0.0	0.0
Verrucomicrobia	0.6	0.4	0.5	0.3	0.3	0.2	0.1	0.2	0.0	0.4	0.2	0.2	0.1	0.0	0.0

Table S3.1: Continued.

#OTU ID	CT1_4	CT2_4	CT3_4	CT1_7	CT2_7	CT3_7	SC1_4	SC2_4	SC3_4	SC1_7	SC2_7	SC3_7
OTU131;Deltotrichonympha_nana	2.0	2.6	2.7	0.9	3.8	0.1	1.8	1.9	3.4	1.0	0.4	0.2
OTU77;Koruga_bonita	0.2	0.2	0.2	0.1	0.4	0.0	0.1	0.1	0.3	0.2	0.0	0.0
OTU161;Metadevescovina_extranea	0.1	0.3	0.1	0.0	0.2	0.0	0.2	0.1	0.1	0.2	0.5	0.1
Parabasalia	2.3	3.2	3.0	1.0	4.3	0.2	2.1	2.1	3.7	1.4	1.0	0.2
OTU86;f_Methanobacteriaceae	0.3	0.2	0.3	0.1	0.2	0.1	0.5	0.2	0.2	0.3	0.2	0.1
OTU28;g_Methanobacterium	0.2	0.2	0.4	0.0	0.0	0.0	0.3	0.3	0.2	0.1	0.0	0.0
OTU9;g_Methanobacterium; s_beijingense	4.2	1.8	2.3	1.3	1.8	0.9	3.8	2.0	1.8	2.4	1.3	0.8
OTU31;g_Methanobrevibacter	0.3	0.3	0.3	0.0	0.3	0.1	0.6	0.3	0.2	0.5	1.4	0.0
OTU38;g_Methanobrevibacter; s_arboriphilus_1	5.6	4.4	4.9	2.5	1.6	2.2	4.5	3.6	3.2	1.9	1.9	1.4
OTU89;g_Methanobrevibacter; s_arboriphilus_2	0.6	0.3	0.4	0.9	0.5	1.0	0.3	0.6	0.3	0.3	0.4	0.6
Euryarchaeota	11.2	7.2	8.7	4.9	4.4	4.3	9.9	7.1	5.9	5.5	5.2	2.9
OTU113;p_Acidobacteria; o_MVS-40	0.7	0.6	0.7	0.2	0.4	0.1	0.4	0.5	0.5	0.2	0.2	0.5
Acidobacteria	0.7	0.6	0.7	0.2	0.4	0.1	0.4	0.5	0.5	0.2	0.2	0.5
OTU52;g_Corynebacterium	0.0	0.0	0.6	5.1	0.0	2.1	0.0	0.0	0.0	3.0	3.1	1.1
OTU51;g_Microbacterium_2	0.0	0.0	0.1	2.1	0.0	0.9	0.0	0.0	0.0	0.7	0.2	0.1
OTU55;g_Microbacterium_3	0.0	0.0	0.1	0.5	0.0	2.7	0.0	0.0	0.0	0.0	0.2	0.2
OTU147;g_Microbacterium_1	0.0	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.3	0.2	0.1
OTU10;g_Propionicimonas_2	0.1	0.2	0.1	0.2	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0.3
OTU12;g_Propionicimonas_3	0.1	0.3	0.2	0.2	0.2	0.1	0.2	0.2	0.1	0.2	0.3	0.2
OTU53;g_Propionicimonas_1	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1
OTU73;g_Xylanimicrobium	0.0	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.2	0.2
OTU24;f_Propionibacteriaceae	1.2	1.0	1.6	2.4	2.0	2.6	1.0	1.7	0.9	0.8	1.8	3.5
OTU42;g_Tsukamurella	0.0	0.0	0.3	0.4	0.0	7.4	0.0	0.0	0.0	0.6	0.0	0.2
OTU23;f_Coriobacteriaceae_1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.1
OTU37;f_Coriobacteriaceae_2	0.1	0.3	0.2	0.3	0.2	0.2	0.3	0.3	0.2	0.2	0.4	0.2
Actinobacteria	1.8	2.0	3.6	11.7	2.8	16.5	2.1	2.8	1.9	6.3	6.8	6.4
OTU134;o_Bacteroidales_1	0.6	0.2	0.5	0.3	0.1	0.9	0.4	0.3	0.2	0.4	0.9	2.1
OTU17;o_Bacteroidales_4	0.2	0.1	0.2	0.2	0.1	0.2	0.1	0.1	0.1	0.6	0.4	0.2
OTU72;o_Bacteroidales_6	3.0	1.6	1.2	0.7	0.7	0.2	0.6	1.1	1.9	1.9	0.4	0.4
OTU82;o_Bacteroidales_3	1.4	0.8	0.5	0.3	0.4	0.1	0.3	0.5	1.0	0.9	0.2	0.2
OTU114;o_Bacteroidales_7	0.3	0.2	0.1	0.1	0.1	0.0	0.2	0.2	0.1	0.1	0.1	0.1
OTU159;o_Bacteroidales_2	0.3	0.3	0.2	0.1	0.3	0.0	0.3	0.1	0.2	0.1	0.1	0.0
OTU160;o_Bacteroidales_5	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.1	0.1
OTU95;f_Bacteroidaceae	0.3	0.4	0.4	0.3	0.4	0.1	0.4	0.3	0.3	0.2	0.2	0.1
OTU22;g_Bacteroides_2	0.1	0.0	0.0	0.1	0.0	0.1	0.1	0.1	0.2	0.2	0.4	0.4
OTU63;g_Bacteroides_3	0.1	0.0	0.0	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.0
OTU145;g_Bacteroides_1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.3	0.3	0.2	0.4	0.4
OTU20;g_Bacteroides; s_eggerthii	0.2	0.1	0.1	0.6	0.2	0.1	0.3	0.8	0.5	0.8	1.4	0.8
OTU142;g_Bacteroides; s_uniformis	0.1	0.2	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.2
OTU120;o_Bacteroidales; f_p-2534-18B5_2	0.2	0.1	0.1	0.0	0.1	0.0	0.1	0.1	0.2	0.2	0.0	0.0
OTU137;o_Bacteroidales; f_p-2534-18B5_1	1.2	1.0	1.0	0.9	0.9	0.1	0.8	0.7	1.6	2.2	0.5	0.4
OTU49;f_Porphyrimonadaceae	0.1	0.1	0.2	0.3	0.2	0.2	0.4	0.3	0.5	0.3	0.3	0.8
OTU19;g_Candidatus Azobacteroides_4	3.3	2.7	3.4	2.1	2.6	1.6	2.6	2.7	2.9	1.9	2.3	1.9
OTU41;g_Candidatus Azobacteroides_2	0.2	0.2	0.3	0.1	0.1	0.0	0.2	0.2	0.1	0.1	0.1	0.1
OTU111;g_Candidatus Azobacteroides_1	0.2	0.1	0.2	0.2	0.2	0.3	0.1	0.3	0.2	0.1	0.1	0.1
OTU124;g_Candidatus Azobacteroides_5	0.7	0.5	0.7	0.4	0.7	0.2	0.5	0.5	0.6	0.4	0.4	0.4

Table S3.1: Continued.

#OTU ID	CT1_4	CT2_4	CT3_4	CT1_7	CT2_7	CT3_7	SC1_4	SC2_4	SC3_4	SC1_7	SC2_7	SC3_7
OTU128;g__Candidatus Azobacteroides_3	2.0	1.5	2.5	1.5	1.9	0.6	1.5	1.3	1.4	3.4	1.5	0.7
OTU74;g__Dysgonomonas_4	0.1	0.0	0.0	0.0	0.0	1.6	0.1	0.0	0.7	0.1	3.5	1.6
OTU78;g__Dysgonomonas_3	0.5	0.4	0.3	0.1	0.2	0.0	0.6	0.3	0.3	0.1	0.1	0.1
OTU148;g__Dysgonomonas_1	2.6	4.9	5.5	7.3	5.5	6.3	5.1	5.4	6.0	6.0	9.5	6.4
OTU158;g__Dysgonomonas_2	0.2	0.3	0.3	0.5	0.4	0.3	0.2	0.3	0.3	0.2	0.3	0.4
OTU40;g__Parabacteroides	0.3	0.2	0.3	0.3	0.4	0.0	0.6	0.2	0.4	0.2	0.2	0.2
OTU59;g__Parabacteroides	0.2	0.3	0.3	0.4	0.3	0.1	0.3	0.2	0.4	0.2	0.2	0.2
OTU68;g__Parabacteroides	2.4	2.0	2.0	2.7	2.9	0.5	3.1	2.6	3.2	2.3	1.7	1.5
OTU71;g__Parabacteroides	2.2	2.0	1.9	1.3	2.0	0.8	1.7	1.9	1.7	1.4	1.3	1.2
OTU85;g__Parabacteroides	0.0	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.0
OTU88;g__Parabacteroides	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
OTU98;g__Parabacteroides	0.1	0.1	0.1	0.0	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.1
OTU99;g__Parabacteroides	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0
OTU115;g__Parabacteroides	0.2	0.1	0.1	0.2	0.2	0.0	0.2	0.2	0.2	0.1	0.1	0.1
OTU126;g__Parabacteroides	0.5	0.5	0.7	0.4	0.7	0.2	0.5	0.5	0.5	0.4	0.4	0.4
OTU139;g__Parabacteroides	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
OTU67;g__Tannerella	3.2	1.4	2.0	1.6	2.3	0.6	2.8	2.1	2.7	1.0	1.9	1.1
OTU75;g__Tannerella	0.5	0.3	0.6	0.1	0.0	0.0	0.4	0.3	0.2	0.1	0.0	0.0
OTU143;g__Tannerella	0.8	0.3	0.5	0.1	0.4	0.1	0.5	0.2	0.4	0.1	0.2	0.1
OTU125;g__Prevotella	0.5	0.2	0.3	0.4	0.5	0.0	0.3	0.3	0.4	0.4	0.1	0.1
OTU32;o__Bacteroidales; f__RF16_1	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.2	0.1	0.2	0.1	0.1
OTU90;o__Bacteroidales; f__RF16_2	0.3	0.3	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.2	0.1	0.0
OTU153;o__Bacteroidales; f__RF16_3	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.2	0.1
OTU132;o__Bacteroidales; f__S24-7	0.2	0.2	0.4	3.4	0.6	3.3	0.3	0.4	0.3	0.7	1.5	1.0
OTU61;g__Chryseobacterium	0.0	0.0	0.0	0.7	0.2	0.0	0.1	0.0	0.0	2.0	0.0	0.0
OTU57;g__Blattabacterium	3.9	20.5	11.5	3.6	8.8	20.2	8.1	13.9	4.0	2.8	2.6	25.4
OTU84;g__Blattabacterium	0.2	0.4	0.3	0.1	0.2	0.2	0.2	0.4	0.1	0.1	0.1	0.3
OTU140;g__Blattabacterium	0.1	0.2	0.1	0.0	0.1	0.2	0.1	0.2	0.1	0.1	0.0	0.2
OTU151;g__Blattabacterium	0.1	0.1	0.1	0.1	0.1	0.3	0.1	0.2	0.1	0.0	0.0	0.3
OTU155;g__Blattabacterium	0.1	0.4	0.3	0.1	0.1	0.2	0.2	0.4	0.1	0.1	0.1	0.4
OTU60;g__Sphingobacterium; s__multivorum	0.0	0.0	0.3	5.2	0.0	0.1	0.0	0.0	0.0	0.1	0.1	0.1
Bacteroidetes	34.3	46.0	40.0	37.2	36.0	40.2	35.1	40.4	35.3	33.7	34.4	51.1
OTU26;c__Endomicrobia_5	0.3	0.4	0.3	0.1	0.4	0.0	0.4	0.3	0.3	0.1	0.1	0.0
OTU27;c__Endomicrobia_3	14.8	7.7	9.3	1.6	10.0	0.2	8.6	7.5	10.1	6.0	0.9	0.6
OTU81;c__Endomicrobia_2	0.2	0.2	0.2	0.1	0.2	0.0	0.2	0.2	0.2	0.2	0.0	0.0
OTU92;c__Endomicrobia_4	0.3	0.2	0.2	0.2	0.4	0.0	0.3	0.2	0.2	0.2	0.1	0.0
OTU102;c__Endomicrobia_7	0.5	0.3	0.5	0.0	0.5	0.0	0.8	0.3	0.6	0.2	0.1	0.0
OTU105;c__Endomicrobia_8	0.3	0.2	0.2	0.1	0.1	0.0	0.2	0.2	0.2	0.2	0.0	0.0
OTU122;c__Endomicrobia_1	0.9	0.6	0.6	0.0	0.0	0.0	0.5	0.4	0.5	0.2	0.0	0.0
OTU144;c__Endomicrobia_6	1.9	0.4	0.7	0.1	0.8	0.0	0.9	0.6	1.4	0.2	0.4	0.1
Elusimicrobia	19.2	9.8	12.1	2.2	12.4	0.4	12.0	9.7	13.5	7.2	1.6	0.8
OTU36;g__Enterococcus	0.4	0.6	0.3	0.9	0.4	0.6	0.3	0.2	0.5	0.2	0.2	1.0
OTU29;f__Leuconostocaceae_1	0.9	3.7	1.9	1.4	4.4	1.5	2.8	2.2	1.9	4.0	2.9	1.3
OTU70;f__Leuconostocaceae_2	9.4	5.7	8.1	9.9	5.1	7.1	14.3	13.0	11.7	19.1	18.5	7.2
OTU100;f__Leuconostocaceae_3	0.5	2.5	1.2	1.0	2.9	1.2	1.8	1.5	1.2	2.4	1.7	0.9
OTU14;g__Lactococcus_4	0.5	0.4	0.3	1.5	1.9	0.6	0.1	0.1	0.1	0.3	0.6	1.0

Table S3.1: Continued.

#OTU ID	CT1_4	CT2_4	CT3_4	CT1_7	CT2_7	CT3_7	SC1_4	SC2_4	SC3_4	SC1_7	SC2_7	SC3_7
OTU35;g__Lactococcus_2	0.0	0.0	0.0	0.1	0.1	0.2	0.0	0.0	0.0	0.0	0.1	0.0
OTU50;g__Lactococcus_3	0.0	0.1	0.1	0.0	0.1	0.4	0.1	0.0	0.0	0.2	0.2	0.0
OTU108;g__Lactococcus_1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.1	0.0
OTU80;o__Clostridiales_4	0.1	0.1	0.1	0.1	0.2	0.0	0.1	0.1	0.1	0.1	0.0	0.0
OTU106;o__Clostridiales_3	0.5	0.3	0.5	0.3	0.2	0.4	0.6	0.5	0.3	1.1	0.8	0.4
OTU123;o__Clostridiales_1	0.0	0.5	0.1	0.0	0.0	0.0	0.1	0.2	0.1	0.1	0.0	0.1
OTU138;o__Clostridiales_2	0.1	0.1	0.0	0.2	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.0
OTU141;o__Clostridiales_5	0.1	0.3	0.2	0.1	0.3	0.0	0.2	0.2	0.3	0.2	0.0	0.1
OTU18;f__Mogibacteriaceae_7	0.1	0.3	0.2	0.3	0.1	0.2	0.2	0.3	0.1	0.3	0.4	0.2
OTU25;f__Mogibacteriaceae_3	1.0	1.6	1.7	1.6	1.2	0.8	1.5	1.8	1.0	1.6	1.5	1.0
OTU45;f__Mogibacteriaceae_2	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.0
OTU47;f__Mogibacteriaceae_9	0.1	0.1	0.2	0.0	0.2	0.2	0.2	0.3	0.4	0.2	0.3	0.1
OTU94;f__Mogibacteriaceae_5	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.1
OTU101;f__Mogibacteriaceae_8	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.1
OTU129;f__Mogibacteriaceae_4	0.4	0.2	0.3	0.2	0.2	0.1	0.3	0.5	0.3	0.3	0.2	0.2
OTU150;f__Mogibacteriaceae_1	0.3	0.3	0.3	0.2	0.2	0.0	0.2	0.4	0.3	0.1	0.1	0.1
OTU154;f__Mogibacteriaceae_6	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1
OTU112;f__Clostridiaceae_2	0.4	0.3	0.2	1.5	0.5	1.4	0.3	0.5	0.2	1.0	0.5	0.6
OTU118;f__Clostridiaceae_1	0.1	0.0	0.1	0.2	0.1	0.2	0.0	0.1	0.2	0.1	0.1	0.2
OTU6;g__Clostridium; s__subterminale	0.0	0.0	0.0	0.2	0.0	0.4	0.0	0.0	0.0	0.1	0.1	0.4
OTU109;g__Anaerofustis_1	0.1	0.4	0.1	0.0	0.2	0.0	0.1	0.2	0.1	0.0	0.1	0.1
OTU110;g__Anaerofustis_2	0.2	0.2	0.1	0.0	0.1	0.0	0.1	0.2	0.1	0.0	0.0	0.2
OTU8;f__Lachnospiraceae_1	0.0	0.0	0.0	0.2	0.0	0.1	0.0	0.0	0.1	0.1	0.1	0.3
OTU34;f__Lachnospiraceae_2	0.2	0.0	0.0	0.4	0.0	0.7	0.0	0.0	0.0	0.1	0.5	0.3
OTU69;f__Lachnospiraceae_3	0.5	0.2	1.1	7.1	1.9	4.9	0.7	1.3	4.3	4.2	8.3	11.3
OTU56;g__Anaerostipes	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
OTU11;f__Peptostreptococcaceae_1	0.1	0.0	0.1	0.1	0.1	0.3	0.0	0.1	0.2	0.1	0.1	0.2
OTU48;f__Peptostreptococcaceae_5	0.2	0.1	0.3	0.3	0.3	0.7	0.1	0.1	0.3	0.2	0.2	0.4
OTU116;f__Peptostreptococcaceae_4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
OTU121;f__Peptostreptococcaceae_2	0.1	0.0	0.1	0.4	0.1	0.4	0.0	0.0	0.1	0.1	0.2	0.2
OTU136;f__Peptostreptococcaceae_3	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
OTU54;f__Ruminococcaceae_7	0.2	0.1	0.1	0.2	0.1	0.1	0.0	0.1	0.1	1.0	0.3	0.1
OTU93;f__Ruminococcaceae_2	0.1	0.3	0.2	0.2	0.3	0.0	0.2	0.2	0.1	0.0	0.1	0.2
OTU103;f__Ruminococcaceae_3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
OTU107;f__Ruminococcaceae_8	0.3	0.5	0.3	0.0	0.1	0.0	0.6	0.2	0.3	0.3	0.1	0.0
OTU130;f__Ruminococcaceae_1	0.2	0.1	0.5	0.1	0.2	0.0	0.2	0.1	0.2	0.3	0.1	0.1
OTU133;f__Ruminococcaceae_5	0.1	0.2	0.1	0.0	0.1	0.0	0.3	0.2	0.2	0.0	0.0	0.0
OTU149;f__Ruminococcaceae_4	0.1	0.2	0.1	0.1	0.2	0.1	0.1	0.2	0.1	0.1	0.1	0.2
OTU156;f__Ruminococcaceae_6	0.1	0.2	0.1	0.0	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.0
OTU46;g__Phascolarctobacterium	0.1	0.0	0.1	0.1	0.1	0.5	0.0	0.0	0.1	0.2	0.3	0.2
OTU21;f__Erysipelotrichaceae	0.2	0.1	0.3	0.9	0.2	3.5	0.1	0.2	0.1	0.3	1.1	1.3
Firmicutes	18.4	20.5	20.1	30.6	22.7	27.3	27.1	26.2	26.3	39.1	41.0	30.3
OTU96;g__Fusobacterium_2	0.1	0.2	0.1	0.2	0.2	0.1	0.2	0.2	0.2	0.1	0.2	0.1
OTU97;g__Fusobacterium_1	2.3	3.7	3.6	3.8	4.9	1.5	3.7	3.5	5.8	1.6	3.0	2.6
OTU119;g__Fusobacterium_3	0.1	0.2	0.2	0.2	0.2	0.0	0.2	0.1	0.3	0.1	0.1	0.1
Fusobacteria	2.6	4.0	3.9	4.1	5.3	1.7	4.1	3.8	6.3	1.8	3.3	2.8

Table S3.1: Continued.

#OTU ID	CT1_4	CT2_4	CT3_4	CT1_7	CT2_7	CT3_7	SC1_4	SC2_4	SC3_4	SC1_7	SC2_7	SC3_7
OTU66;f__Pirellulaceae	0.2	0.3	0.4	0.2	0.3	0.2	0.3	0.3	0.2	0.1	0.2	0.3
Planctomycetes	0.2	0.3	0.4	0.2	0.3	0.2	0.3	0.3	0.2	0.1	0.2	0.3
OTU79;o__Rickettsiales	0.3	0.3	0.0	0.0	4.9	0.0	0.1	0.3	0.0	0.0	0.0	0.0
OTU43;c__Betaproteobacteria	0.2	0.5	0.4	0.2	0.3	0.1	0.4	0.3	0.4	0.1	0.2	0.3
OTU65;f__Oxalobacteraceae	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.2	0.2
OTU13;f__Rhodocyclaceae; g__TS34_2	1.5	1.0	1.6	0.6	1.4	0.3	1.1	2.4	1.1	0.5	0.7	1.5
OTU135;f__Rhodocyclaceae; g__TS34_1	0.1	0.1	0.2	0.0	0.1	0.0	0.1	0.2	0.1	0.0	0.0	0.1
OTU39;f__Enterobacteriaceae_1	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.1	0.0
OTU64;f__Enterobacteriaceae_2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0
OTU76;f__Enterobacteriaceae_3	0.4	0.0	0.0	0.0	0.2	4.6	0.5	0.1	0.0	0.1	0.8	0.0
OTU44;g__Acinetobacter	0.0	0.0	0.3	1.2	0.1	0.1	0.0	0.1	0.0	0.9	0.3	0.2
OTU33;g__Pseudomonas	0.0	0.0	0.0	0.4	0.0	0.6	0.0	0.0	0.0	0.0	1.0	0.1
OTU4;g__Stenotrophomonas; s__geniculata	0.0	0.0	0.1	1.0	0.0	0.4	0.0	0.0	0.0	0.1	0.2	0.0
Proteobacteria	2.8	2.0	2.7	3.8	7.1	6.5	2.4	3.4	1.7	2.0	3.6	2.4
OTU1;g__Treponema_7	0.6	0.3	0.3	0.1	0.4	0.0	0.4	0.3	0.3	0.3	0.1	0.0
OTU2;g__Treponema_6	1.4	0.5	0.8	0.1	0.5	0.0	0.8	0.6	0.9	0.1	0.5	0.1
OTU5;g__Treponema_1	0.2	0.1	0.2	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.1
OTU7;g__Treponema_10	0.2	0.1	0.1	0.1	0.1	0.0	0.1	0.2	0.2	0.1	0.1	0.1
OTU15;g__Treponema_11	0.2	0.1	0.1	0.1	0.1	0.0	0.2	0.2	0.1	0.1	0.0	0.1
OTU58;g__Treponema_9	0.6	0.3	0.2	0.0	0.1	0.0	0.2	0.0	0.3	0.1	0.0	0.0
OTU62;g__Treponema_2	0.4	0.1	0.1	0.0	0.2	0.0	0.3	0.2	0.2	0.1	0.1	0.0
OTU91;g__Treponema_3	0.2	0.1	0.1	0.1	0.1	0.0	0.2	0.1	0.1	0.1	0.1	0.0
OTU104;g__Treponema_8	0.5	0.4	0.3	0.1	0.3	0.0	0.3	0.3	0.3	0.2	0.1	0.2
OTU146;g__Treponema_4	0.3	0.1	0.1	0.1	0.1	0.0	0.2	0.2	0.1	0.1	0.1	0.1
OTU157;g__Treponema_5	0.1	0.1	0.2	0.0	0.1	0.0	0.2	0.1	0.1	0.1	0.1	0.1
OTU152;g__Treponema; s__primitia	0.2	0.2	0.1	0.1	0.1	0.0	0.2	0.2	0.1	0.1	0.1	0.0
OTU3;g__Treponema; s__sp5	0.4	0.1	0.2	0.1	0.1	0.0	0.2	0.1	0.1	0.0	0.1	0.1
Spirochaetes	5.2	2.8	2.8	1.0	2.3	0.3	3.5	2.6	2.8	1.3	1.4	0.9
OTU16;f__Dethiosulfovibrionaceae; g__TG5_2	0.1	0.1	0.1	0.0	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.0
OTU87;f__Dethiosulfovibrionaceae; g__TG5_1	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.1
OTU30;g__Candidatus Tammella	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1
OTU127;g__Candidatus Tammella	0.1	0.1	0.1	0.0	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0
Synergistetes	0.4	0.2	0.3	0.2	0.3	0.1	0.3	0.3	0.2	0.1	0.2	0.2
OTU117;c__Mollicutes; o__RsaHF231	0.7	0.6	1.5	2.9	1.4	2.3	0.4	0.5	1.1	0.9	1.1	1.1
Tenericutes	0.7	0.6	1.5	2.9	1.4	2.3	0.4	0.5	1.1	0.9	1.1	1.1
OTU83;f__Cerasioccaceae	0.4	0.7	0.1	0.0	0.2	0.0	0.4	0.3	0.4	0.4	0.1	0.0
Verrucomicrobia	0.4	0.7	0.1	0.0	0.2	0.0	0.4	0.3	0.4	0.4	0.1	0.0

ER1-3_0 *E. regnans* at Ohr CC1-3_4 *C. citrodora* at Day 4 CT1-3_7 Cotton at Day 7
ER1-3_4 *E. regnans* at Day 4 CC1-3_7 *C. citrodora* at Day 7 SC1-3_4 Sugarcane mulch at Day 4
ER1-3_7 *E. regnans* at Day 7 CT1-3_4 Cotton at Day 4 SC1-3_7 Sugarcane mulch at Day 7

Table S3.2: Relative abundances ($\geq 0.2\%$) of gut profiles from three same colonies using different sequencing platforms.

Bacterial phylum	<i>Nasutitermes corniger</i>		<i>Trinervitermes sp</i>		<i>Cubitermes ugandensis</i>	
	Pyrotag [^]	Itag [*]	Pyrotag [^]	Itag [*]	Pyrotag [^]	Itag [*]
Acidobacteria	0.3	1.3	0.8	1.1	0.0	0.3
Actinobacteria	0.8	1.0	2.2	1.8	2.8	3.8
Armatimonadetes	0.0	0.0	0.0	0.0	0.0	0.0
Bacteroidetes	6.4	4.9	14.7	17.3	21.2	8.8
Candidate phylum BD1 5	0.0	0.0	0.0	0.0	0.0	0.0
Candidate phylum BRC1	0.0	0.0	0.0	0.0	0.0	0.0
Candidate phylum OP11	0.0	0.3	0.1	0.1	0.0	0.4
Candidate phylum SR1	0.0	0.3	0.1	0.1	0.1	0.0
Candidate phylum TG3	13.0	3.9	2.0	1.5	1.4	0.0
Candidate phylum TM7	1.0	0.9	0.5	0.6	0.9	0.5
Chlamydiae	0.0	0.0	0.0	0.0	0.0	0.0
Chlorobi	0.0	0.0	0.0	0.0	0.1	0.2
Chloroflexi	0.0	0.0	0.0	0.0	0.1	0.1
Cyanobacteria	0.1	0.4	0.0	0.0	0.1	0.0
Deferribacteres	0.0	0.0	0.1	0.1	0.0	0.1
Deinococcus-Thermus	0.0	0.0	0.0	0.0	0.0	0.0
Elusimicrobia	0.0	0.0	0.0	0.0	0.3	0.6
Fibrobacteres	10.0	9.1	0.1	0.1	0.1	0.0
Firmicutes	6.7	6.7	24.1	23.8	52.3	55.5
Fusobacteria	0.1	0.2	0.0	0.0	0.1	0.3
Gemmatimonadetes	0.0	0.0	0.0	0.0	0.0	0.0
Lentisphaerae	0.0	0.1	0.1	0.1	0.3	0.9
NPL-UPA2	0.0	0.0	0.0	0.0	0.0	0.0
Nitrospirae	0.0	0.0	0.0	0.0	0.0	0.0
Planctomycetes	0.2	0.3	0.0	0.0	0.4	0.6
Proteobacteria	4.0	3.1	4.2	3.8	6.4	11.9
Spirochaetes	55.5	66.2	49.8	47.9	10.7	11.5
Synergistetes	0.1	0.7	1.0	1.1	0.5	2.4
Tenericutes	-	0.1	-	0.0	-	0.0
Verrucomicrobia	0.0	0.2	0.0	0.0	0.1	0.8
WCHB1-60	0.0	0.0	0.0	0.0	0.0	0.0
unknown	1.6	0.7	0.2	0.4	1.8	1.6

[^] Dietrich et al., 2014

^{*} Mikaelyan et al., 2015

Appendix C: Supplementary figures and tables for Chapter 4

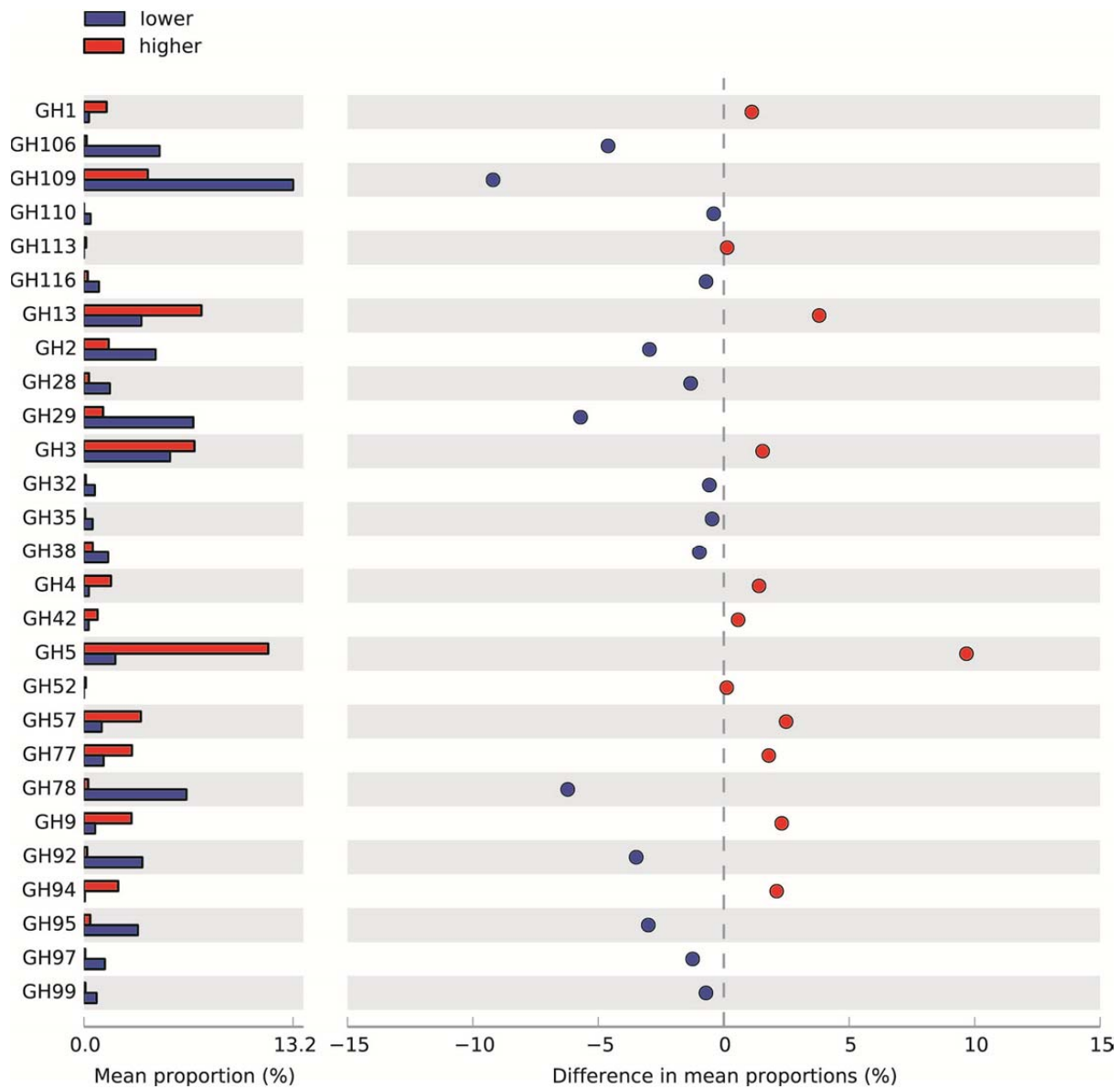


Figure S4.1: Glycoside hydrolases (GHs) encoding genes with significant difference in mean proportion between lower and higher termites.

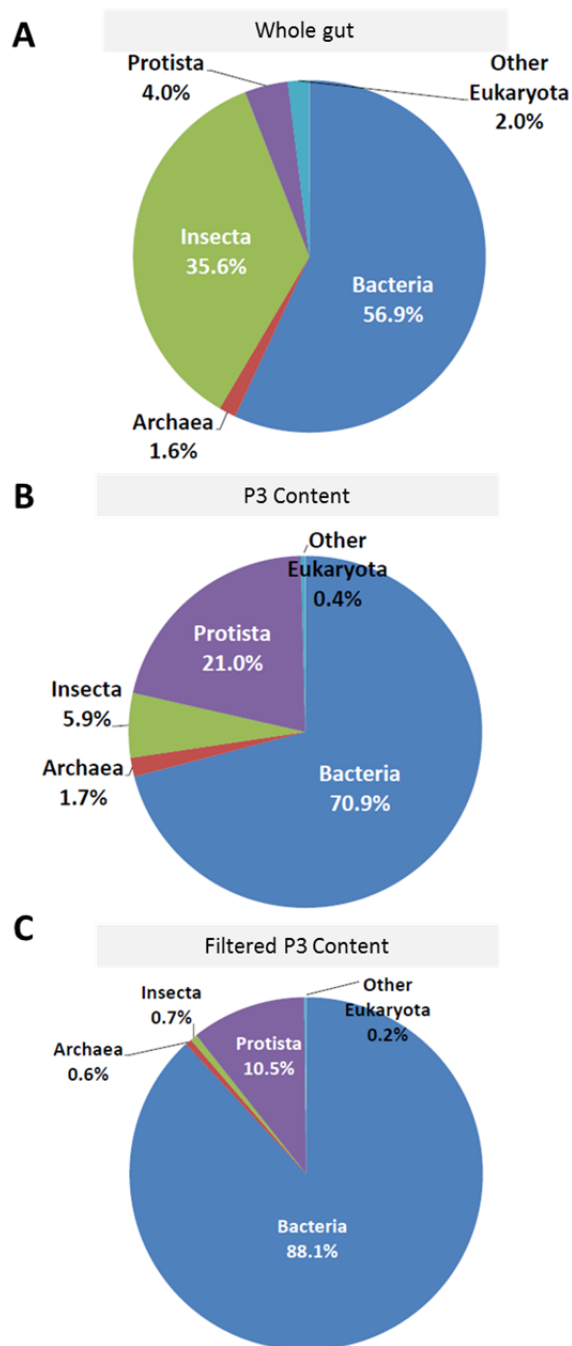


Figure S4.2: Distribution of metagenomic reads derived from termite host and microbial symbionts in three different gut DNAs extracted from A) whole gut sample, B) P3 gut content, C) filtered P3 gut content.

Table S4.1: Sequencing, assembly, annotation and binning statistics of metagenomes.

Metagenome	Samples	Sequencing platform	Sequencing			Assembly				Annotation		Binning
			No. of Illumina reads X2 (paired-end)	No. of QC reads	No. of QC bp	No. of bases went into contigs (bp)	No. of contigs*	Highest contig	N50	No. of coding bases	Percent of COG (%)	Population genomes
M	IN012001	HiSeq	62,575,107	95,378,037	10,859,265,807	324,991,198	248,238	164,466	1,432	8,169,273	21.8	65 [#]
N	MC05	(2 X 100 bp)	84,465,475	121,159,426	12,287,121,832	326,777,528	298,626	110,183	1,070	10,597,450	14.6	101 [#]
P	TN01	MiSeq (2 X 250 bp)	11,821,118	12,212,753	3,484,459,158	262,533,523	233,941	78,809	1,142	2,136,343	17.8	13
MD	Whole gut1	HiSeq	19,680,253	36,672,375	3,455,608,272	85,903,917	69,432	142,742	1,391	1,208,144	12.9	-
MT	TV022002	(2 X 100 bp)	31,524,589	58,589,923	5,522,825,295	147,868,612	162,401	108,154	851	2,233,301	8.4	-

*Longer than 500bp

[#] Differential coverage binning is based on co-assembly (see Chapter 5 for detailed analysis)

Table S4.2: Glycoside hydrolases (GHs) identified in the metagenomes, organised by functional category.

Family	<i>Microcerotermes</i>		<i>Nasutitermes</i>		<i>Porotermes</i>		<i>Mastotermes_DW</i>		<i>Mastotermes_TV</i>	
	(%)	Gene count*	(%)	Gene count*	(%)	Gene count*	(%)	Gene count*	(%)	Gene count*
Cellulase										
<i>Endocellulases (active on internal β-1,4 glucosidic bonds) EC 3.2.1.4</i>										
GH5	11.5	7796	11.8	6705	1.9	204	2.5	165	1.5	103
GH9	3.1	2121	2.9	1642	0.4	47	1.1	74	0.5	36
GH44	0.0	7	0.1	27	-	-	-	-	-	-
GH45	0.7	473	1.0	573	-	-	-	-	-	-
<i>β-glucosidases (producing glucose from cellobiose) EC3.2.1.21</i>										
GH1	1.5	1005	1.4	784	0.5	56	-	-	0.4	28
GH3	6.9	4632	7.1	4062	5.2	554	5.3	349	5.9	392
GH4	1.5	1039	1.9	1076	0.6	64	-	-	0.3	22
GH16	0.8	568	0.4	204	1.1	121	1.1	70	1.0	66
GH17	-	-	-	-	0.1	6	0.2	14	-	-
<i>Cellobiose/chitobiose phosphorylase</i>										
GH84	-	-	0.0	8	-	-	-	-	-	-
GH94	2.2	1460	2.2	1238	0.2	19	-	-	-	-
Subtotal	28.3	19100	28.7	16319	10.0	1070	10.2	672	9.6	646
Hemicellulases										
<i>Depolymerising</i>										
<i>Endohemicellulase EC 3.2.1.8</i>										
GH8	2.1	1391	0.5	293	0.1	6	-	-	-	-
GH10	8.4	5700	4.5	2587	0.6	63	0.5	32	1.0	67
GH11	3.1	2082	2.6	1498	-	-	-	-	-	-
GH26	1.7	1170	0.2	130	0.4	44	0.5	32	0.2	11
GH28	0.3	216	0.3	177	1.5	159	1.8	118	1.6	109
GH53	1.4	935	0.9	498	0.1	8	0.3	18	0.2	12
GH113	0.1	96	0.1	72	-	-	-	-	-	-
<i>Exohemicellulase EC 3.2.1.37/EC 3.2.1.25</i>										
GH39	1.5	1023	0.9	538	0.6	59	0.2	10	0.1	5
GH52	0.1	86	0.1	63	-	-	-	-	-	-
GH2	1.9	1252	1.3	730	3.1	334	5.5	364	4.9	330
GH120	0.0	21	1.3	709	0.7	73	0.5	33	0.3	23
GH35	0.1	55	0.1	43	0.4	42	0.8	50	0.5	32
GH42	0.8	549	0.9	521	0.5	49	0.2	13	0.2	15
GH116	0.2	139	0.3	157	0.7	74	1.2	78	1.0	64
<i>Debranching</i>										
<i>α-glucuronidases EC 3.2.1.139</i>										
GH67	0.3	201	0.1	71	0.3	33	0.2	12	0.1	6
<i>α-arabinofuranosidases EC 3.2.1.55</i>										
GH43	6.0	4061	6.9	3935	4.8	514	5.8	379	5.9	392
GH51	0.9	623	1.2	669	1.4	154	1.6	104	1.4	90
GH54	-	-	0.0	4	-	-	-	-	-	-

Table S4.2: Continued.

Family	Microcerotermes		Nasutitermes		Porotermes		Mastotermes_DW		Mastotermes_TV	
	(%)	Gene count*	(%)	Gene count*	(%)	Gene count*	(%)	Gene count*	(%)	Gene count*
<i>α-galactosidases EC 3.2.1.22</i>						0				
GH4	1.5	1039	1.9	1076	0.6	64	-	-	0.3	22
GH27	0.1	31	0.9	504	0.3	29	0.8	52	0.5	33
GH36	0.3	194	0.2	117	0.4	47	1.1	71	0.5	33
GH57	3.5	2356	3.7	2116	1.3	135	0.8	51	1.3	87
GH110	0.0	7	0.0	7	0.6	58	0.3	21	0.4	25
GH97	0.1	73	0.1	27	1.4	151	1.5	96	1.1	72
<i>Other debranching</i>						0				
GH16	0.8	568	0.4	204	1.1	121	1.1	70	1.0	66
GH29	0.7	472	1.7	982	9.5	1017	5.8	383	5.4	362
GH74	0.1	84	0.0	8	-	-	-	-	-	-
GH31	0.8	551	2.3	1333	0.9	96	0.7	48	0.9	62
GH38	0.5	309	0.6	367	1.5	161	1.7	112	1.4	91
GH76	0.0	21	0.0	6	0.2	17	0.1	7	0.3	20
GH78	0.2	101	0.4	219	7.6	809	5.4	353	6.5	436
GH95	0.4	295	0.4	208	3.9	420	2.7	174	3.6	244
GH106	0.2	107	0.2	103	4.7	498	4.1	272	5.6	372
GH115	1.2	806	0.9	509	0.4	46	0.5	35	0.7	44
Subtotal	39.4	26614	35.9	20478	49.5	5281	45.4	2989	46.7	3127
<i>Chitinase</i>										
GH18	1.7	1140	2.1	1211	0.6	60	0.8	50	1.0	64
GH19	0.1	38	0.0	12	0.0	2	-	-	-	-
GH20	1.1	717	2.1	1220	1.9	201	1.8	120	1.2	79
Subtotal	2.8	1895	4.3	2443	2.5	263	2.6	170	2.1	143
<i>Others</i>										
GH7	-	-	-	-	0.0	4	-	-	-	-
GH13	7.1	4772	7.8	4441	3.9	417	3.3	216	3.7	248
GH15	0.0	6	-	-	0.2	25	-	-	0.1	4
GH22	-	-	-	-	0.0	3	-	-	-	-
GH23	6.5	4379	7.7	4381	4.4	473	4.1	267	6.6	441
GH24	0.3	171	0.1	78	1.0	111	1.4	93	-	-
GH25	0.1	80	0.3	167	1.2	129	1.2	80	0.2	13
GH30	5.4	3654	1.9	1094	0.8	81	1.0	68	0.8	54
GH32	0.1	86	0.1	58	0.8	86	0.6	38	0.7	45
GH33	0.1	68	0.4	203	0.1	11	0.6	39	0.5	32
GH37	-	-	-	-	0.0	2	0.2	12	-	-
GH50	0.0	4	0.0	5	0.1	10	0.3	20	0.3	22
GH55	0.0	22	-	-	0.0	4	0.1	7	0.1	5
GH65	0.3	187	0.2	94	0.6	61	0.2	10	-	-
GH66	0.0	4	-	-	0.0	3	-	-	-	-
GH73	1.9	1288	2.3	1318	3.3	347	2.9	193	2.3	155
GH75	-	-	0.0	8	-	-	-	-	-	-
GH77	2.8	1866	3.3	1888	0.8	83	1.3	86	1.6	110

Table S4.2: Continued.

Family	<i>Microcerotermes</i>		<i>Nasutitermes</i>		<i>Porotermes</i>		<i>Mastotermes_DW</i>		<i>Mastotermes_TV</i>	
	(%)	Gene count*	(%)	Gene count*	(%)	Gene count*	(%)	Gene count*	(%)	Gene count*
GH85	-	-	-	-	0.1	14	-	-	-	-
GH87	0.0	7	0.0	25	-	-	-	-	-	-
GH88	0.4	281	0.1	70	0.4	46	1.1	72	0.8	55
GH89	0.0	8	0.1	31	0.1	9	0.2	11	0.2	14
GH92	0.2	117	0.2	132	2.7	288	4.3	282	4.1	274
GH93	0.0	14	-	-	-	-	0.2	12	0.1	6
GH98	-	-	-	-	-	-	-	-	0.1	5
GH99	0.1	67	0.1	46	0.9	99	0.8	49	0.7	48
GH101	0.0	7	-	-	-	-	-	-	-	-
GH102	0.0	15	-	-	0.1	11	0.3	23	0.1	9
GH103	0.1	86	0.1	37	0.1	14	0.4	25	0.2	14
GH104	0.0	7	-	-	-	-	-	-	1.5	103
GH105	0.5	330	0.2	137	1.0	109	0.7	44	1.0	68
GH108	0.0	13	0.1	37	0.1	14	-	-	-	-
GH109	3.7	2483	4.4	2502	12.2	1301	14.0	924	13.5	902
GH112	0.1	89	0.2	104	-	-	-	-	-	-
GH114	-	-	-	-	0.1	10	0.1	7	-	-
GH117	-	-	-	-	0.4	47	0.3	22	0.1	9
GH123	0.0	6	0.1	40	0.3	34	0.3	16	-	-
GH125	0.1	35	0.0	17	0.3	33	-	-	0.3	19
GH127	0.5	356	1.9	1085	2.1	226	2.3	149	2.8	184
GH128	0.1	44	0.1	37	0.1	10	0.2	14	0.1	6
GH129	-	-	0.0	8	0.0	2	-	-	-	-
GH130	1.5	992	1.7	938	1.2	123	0.7	46	0.3	23
Subtotal	31.9	21547	33.3	18980	39.8	4241	42.9	2827	42.8	2868
Total GHs		69157		58221		10854		6658		6783
Total genes		8012317		10464945		2101114		1195788		2218313

* Gene count is the number of genes weighted by their individual read depth

Table S4.3: Major functions represented between the metagenomes based on Clusters of Orthologous Groups (COG).

Category	COG ID	Subcategory	Microcerotermes		Nasutitermes		Porotermes		Mastotermes_DW		Mastotermes_TV	
			% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*
ENERGY METABOLISM												
<i>Nitrogen metabolism</i>												
C	COG2710	Nitrogenase molybdenum-iron protein	0.13	2091	0.15	2072	0.12	422	0.11	163	0.18	319
E	COG0347	Nitrogen regulatory protein PII	0.12	1899	0.17	2396	0.11	389	0.25	361	0.22	377
P	COG1348	Nitrogenase subunit NifH (ATPase)	0.09	1458	0.12	1706	0.06	223	0.05	77	0.06	104
<i>Hydrogen metabolism</i>												
C	COG0374	Ni, Fe-hydrogenase I large subunit	0.00	16	0.00	21	0.02	71	0.00	0	0.00	0
C	COG1740	Ni, Fe-hydrogenase I small subunit	0.00	32	0.00	21	0.02	74	0.00	0	0.00	7
C	COG3261	Ni, Fe-hydrogenase III large subunit	0.00	24	0.00	52	0.02	80	0.03	41	0.01	18
C	COG3260	Ni, Fe-hydrogenase III small subunit	0.00	24	0.00	31	0.02	86	0.01	16	0.01	22
C	COG4074	H ₂ -forming N ₅ , N ₁₀ -methylene tetrahydropterin dehydrogenase	0.00	0	0.00	0	0.01	42	0.00	0	0.00	7
R	COG4624	Iron only hydrogenase large subunit	0.05	801	0.08	1162	0.02	82	0.04	65	0.07	126
<i>Homoacetogenesis</i>												
C	COG1456	CO dehydrogenase/acetyl-CoA synthase gamma subunit (corrinoid Fe-S protein)	0.02	401	0.01	115	0.01	29	0.00	0	0.01	18
C	COG1152	CO dehydrogenase/acetyl-CoA synthase alpha subunit	0.01	144	0.00	10	0.00	8	0.00	4	0.01	20
C	COG1151	6Fe-6S prismane cluster-containing protein	0.00	72	0.01	209	0.01	38	0.03	43	0.01	18
C	COG2069	CO dehydrogenase/acetyl-CoA synthase delta subunit (corrinoid Fe-S protein)	0.01	168	0.01	73	0.00	15	0.00	6	0.01	18
C	COG1614	CO dehydrogenase/acetyl-CoA synthase beta subunit	0.01	96	0.01	126	0.01	32	0.01	19	0.01	18
C	COG0243	Anaerobic selenocysteine-containing dehydrogenases	0.00	0	0.00	21	0.15	508	0.02	29	0.02	29
C	COG1882	Pyruvate-formate lyase	0.01	104	0.02	241	0.03	95	0.05	72	0.03	53
<i>Pyruvate decarboxylation</i>												
C	COG0280	Phosphotransacetylase	0.06	986	0.06	806	0.12	410	0.09	130	0.15	255
AMINO ACID METABOLISM												
<i>Amino acid transport</i>												
E	COG3842	ABC-type spermidine/putrescine transport systems	0.06	1010	0.09	1308	0.01	46	0.02	36	0.01	16
E	COG4166	ABC-type oligopeptide transport system	0.06	921	0.08	1130	0.03	101	0.01	19	0.03	53
E	COG1176	ABC-type spermidine/putrescine transport system	0.06	1050	0.07	973	0.01	46	0.02	30	0.04	67
E	COG0601	ABC-type dipeptide/oligopeptide/nickel transport systems	0.13	2131	0.19	2689	0.16	555	0.12	166	0.05	84
E	COG4608	ABC-type oligopeptide transport system	0.11	1723	0.12	1695	0.09	300	0.03	50	0.03	51
E	COG1173	ABC-type dipeptide/oligopeptide/nickel transport systems	0.12	2027	0.16	2292	0.13	452	0.10	148	0.03	60
E	COG1115	Na ⁺ /alanine symporter	0.02	353	0.04	576	0.02	86	0.03	37	0.03	55
E	COG0834	ABC-type amino acid transport/signal transduction systems	0.12	1979	0.13	1790	0.13	433	0.03	47	0.03	58
<i>Amino acid and amine oxidation</i>												
C	COG1012	NAD-dependent aldehyde dehydrogenases	0.04	609	0.06	837	0.11	393	0.13	190	0.09	160
P	COG3546	Mn-containing catalase	0.00	0	0.00	42	0.00	2	0.00	6	0.00	0
P	COG0753	Catalase	0.00	32	0.00	10	0.03	107	0.01	20	0.01	13
E	COG0665	Glycine/D-amino acid oxidases (deaminating)	0.00	72	0.00	52	0.02	63	0.00	0	0.01	20
C	COG2414	Aldehyde:ferredoxin oxidoreductase	0.00	72	0.00	10	0.02	53	0.00	0	0.00	0

Table S4.3: Continued.

Category	COG ID	Subcategory	Microcerotermes		Nasutitermes		Porotermes		Mastotermes_DW		Mastotermes_TV	
			% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*
<i>Ammonia transport and assimilation</i>												
P	COG0004	Ammonia permease	0.03	513	0.04	586	0.06	212	0.09	127	0.11	197
E	COG0069	Glutamate synthase domain 2	0.03	489	0.05	649	0.06	210	0.08	110	0.10	177
E	COG0174	Glutamine synthetase	0.04	689	0.01	105	0.04	134	0.05	68	0.08	146
<i>Urease</i>												
E	COG0804	Urea amidohydrolase (urease) alpha subunit	0.01	96	0.02	241	0.01	46	0.03	44	0.01	13
O	COG2371	Urease accessory protein UreE	0.01	144	0.01	126	0.00	2	0.00	0	0.00	0
O	COG0830	Urease accessory protein UreF	0.01	168	0.02	283	0.00	0	0.01	10	0.00	0
E	COG0831	Urea amidohydrolase (urease) gamma subunit	0.01	168	0.02	272	0.01	50	0.03	48	0.01	24
<i>Arginine biosynthesis</i>												
E	COG0002	Acetylglutamate semialdehyde dehydrogenase	0.07	1074	0.06	910	0.09	319	0.10	151	0.10	173
E	COG1364	N-acetylglutamate synthase (N-acetylornithine aminotransferase)	0.06	969	0.05	712	0.06	191	0.06	83	0.09	160
E	COG0137	Argininosuccinate synthase	0.06	1002	0.05	764	0.08	271	0.10	140	0.15	266
<i>Histidine biosynthesis</i>												
E	COG0040	ATP phosphoribosyltransferase	0.06	1010	0.07	1005	0.06	206	0.09	128	0.13	222
E	COG0131	Imidazoleglycerol-phosphate dehydratase	0.08	1242	0.10	1392	0.05	187	0.11	165	0.11	189
E	COG0139	Phosphoribosyl-AMP cyclohydrolase	0.03	553	0.05	733	0.06	200	0.10	143	0.12	209
E	COG1387	Histidinol phosphatase and related hydrolases of the PHP family	0.02	304	0.01	147	0.04	145	0.07	98	0.04	73
<i>Isoleucine biosynthesis</i>												
E	COG0059	Ketol-acid reductoisomerase	0.04	601	0.02	272	0.10	332	0.07	108	0.11	189
E	COG0129	Dihydroxyacid dehydratase/phosphogluconate dehydratase	0.04	689	0.01	209	0.05	183	0.07	108	0.06	111
E	COG0028	Thiamine pyrophosphate-requiring enzymes [acetolactate synthase]	0.09	1482	0.07	1046	0.09	300	0.10	148	0.12	206
E	COG1171	Threonine dehydratase	0.00	32	0.04	513	0.02	63	0.00	0	0.01	11
<i>Leucine biosynthesis</i>												
E	COG0059	Ketol-acid reductoisomerase	0.04	601	0.02	272	0.10	332	0.07	108	0.11	189
C	COG0473	Isocitrate/isopropylmalate dehydrogenase	0.04	657	0.05	659	0.08	271	0.08	116	0.15	255
E	COG0129	Dihydroxyacid dehydratase/phosphogluconate dehydratase	0.04	689	0.01	209	0.05	183	0.07	108	0.06	111
E	COG0066	3-isopropylmalate dehydratase small subunit	0.05	785	0.07	973	0.07	233	0.04	57	0.10	177
E	COG0028	Thiamine pyrophosphate-requiring enzymes [acetolactate synthase]	0.09	1482	0.07	1046	0.09	300	0.10	148	0.12	206
E	COG0065	3-isopropylmalate dehydratase large subunit	0.04	729	0.03	481	0.07	227	0.06	90	0.09	155
<i>Methionine biosynthesis</i>												
E	COG1410	Methionine synthase I	0.04	577	0.06	795	0.07	256	0.05	71	0.03	51
E	COG0626	Cystathionine beta-lyases/cystathionine gamma-synthases	0.07	1138	0.09	1266	0.08	282	0.09	126	0.05	80
E	COG0620	Methionine synthase II (cobalamin-independent)	0.00	48	0.00	10	0.03	111	0.02	31	0.01	11

Table S4.3: Continued.

Category	COG ID	Subcategory	<i>Microcerotermes</i>		<i>Nasutitermes</i>		<i>Porotermes</i>		<i>Mastotermes_DW</i>		<i>Mastotermes_TV</i>	
			% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*
<i>Phenylalanine/tyrosine biosynthesis</i>												
E	COG0169	Shikimate 5-dehydrogenase	0.08	1234	0.15	2135	0.08	265	0.11	155	0.08	142
E	COG0077	Prephenate dehydratase	0.04	697	0.07	963	0.05	158	0.09	134	0.13	231
E	COG0710	3-dehydroquininate dehydratase	0.02	345	0.01	147	0.01	42	0.01	13	0.00	4
<i>Tryptophan biosynthesis</i>												
E	COG0512	Anthranilate/para-aminobenzoate synthases component II	0.08	1250	0.06	858	0.05	162	0.04	62	0.10	171
E	COG0547	Anthranilate phosphoribosyltransferase	0.05	881	0.07	931	0.05	187	0.03	37	0.03	55
E	COG0159	Tryptophan synthase alpha chain	0.07	1074	0.06	858	0.05	185	0.04	53	0.04	71
E	COG0147	Anthranilate/para-aminobenzoate synthases component I	0.06	961	0.04	597	0.05	164	0.04	61	0.04	69
E	COG0169	Shikimate 5-dehydrogenase	0.08	1234	0.15	2135	0.08	265	0.11	155	0.08	142
E	COG0135	Phosphoribosylanthranilate isomerase	0.06	1050	0.09	1203	0.05	174	0.05	72	0.03	44
E	COG0134	Indole-3-glycerol phosphate synthase	0.06	953	0.08	1151	0.05	189	0.04	55	0.02	35
E	COG0710	3-dehydroquininate dehydratase	0.02	345	0.01	147	0.01	42	0.01	13	0.00	4
E	COG0133	Tryptophan synthase beta chain	0.05	793	0.06	879	0.03	99	0.03	38	0.03	58
<i>Valine biosynthesis</i>												
E	COG0059	Ketol-acid reductoisomerase	0.04	601	0.02	272	0.10	332	0.07	108	0.11	189
E	COG0129	Dihydroxyacid dehydratase/phosphogluconate dehydratase	0.04	689	0.01	209	0.05	183	0.07	108	0.06	111
E	COG0436	Aspartate/tyrosine/aromatic aminotransferase	0.26	4158	0.30	4259	0.24	843	0.31	442	0.27	477
E	COG0028	Thiamine pyrophosphate-requiring enzymes [acetolactate synthase	0.09	1482	0.07	1046	0.09	300	0.10	148	0.12	206
<i>Vitamin biosynthesis</i>												
<i>Thiamine</i>												
H	COG1060	Thiamine biosynthesis enzyme ThiH and related uncharacterized enzymes	0.07	1090	0.07	952	0.12	399	0.13	182	0.15	255
H	COG0422	Thiamine biosynthesis protein ThiC	0.02	248	0.01	199	0.09	317	0.07	104	0.03	55
H	COG0352	Thiamine monophosphate synthase	0.04	601	0.06	869	0.07	237	0.07	103	0.11	193
H	COG0476	Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 2	0.10	1546	0.10	1413	0.09	328	0.06	91	0.06	109
H	COG0351	Hydroxymethylpyrimidine/phosphomethylpyrimidine kinase	0.03	489	0.05	753	0.03	105	0.03	39	0.03	49
<i>Biotin</i>												
H	COG0502	Biotin synthase and related enzymes	0.06	1026	0.08	1141	0.04	145	0.05	69	0.02	42
H	COG0161	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase	0.01	160	0.01	84	0.01	44	0.01	10	0.00	4
H	COG0156	7-keto-8-aminopelargonate synthetase and related enzymes	0.03	537	0.02	230	0.09	326	0.09	127	0.14	240
<i>Cobalamin biosynthesis</i>												
P	COG0310	ABC-type Co2+ transport system	0.05	865	0.03	408	0.05	158	0.03	38	0.02	27
H	COG1429	Cobalamin biosynthesis protein CobN and related Mg-chelataes	0.00	0	0.00	0	0.00	0	0.02	23	0.01	18
<i>Inorganic phosphate metabolism</i>												
P	COG1785	Alkaline phosphatase	0.02	361	0.02	314	0.01	40	0.02	32	0.01	22
P	COG1283	Na+/phosphate symporter	0.06	969	0.07	1005	0.03	95	0.03	39	0.03	55
P	COG0855	Polyphosphate kinase	0.06	969	0.05	722	0.03	111	0.03	44	0.02	27

Table S4.3: Continued.

Category	COG ID	Subcategory	Microcerotermes		Nasutitermes		Porotermes		Mastotermes_DW		Mastotermes_TV	
			% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*
CARBOHYDRATE METABOLISM												
<i>Sugar transporter</i>												
G	COG4213	ABC-type xylose transport system	0.08	1362	0.11	1538	0.02	59	0.02	36	0.05	95
G	COG0395	ABC-type sugar transport system	0.50	8189	0.56	7901	0.17	584	0.07	103	0.06	98
G	COG3090	TRAP-type C4-dicarboxylate transport system	0.00	16	0.00	21	0.03	118	0.00	0	0.01	22
G	COG1653	ABC-type sugar transport system	0.11	1859	0.16	2208	0.02	78	0.01	16	0.00	4
G	COG4211	ABC-type glucose/galactose transport system	0.01	232	0.02	220	0.00	13	0.00	4	0.00	0
G	COG2211	Na ⁺ /melibiose symporter and related transporters	0.03	545	0.03	429	0.03	107	0.01	17	0.01	18
G	COG1129	ABC-type sugar transport system	0.08	1290	0.19	2648	0.10	342	0.04	61	0.01	16
G	COG1175	ABC-type sugar transport systems	0.31	4984	0.32	4552	0.08	261	0.03	41	0.04	78
G	COG0738	Fucose permease	0.00	56	0.00	52	0.05	160	0.08	116	0.05	84
G	COG1172	Ribose/xylose/arabinose/galactoside ABC-type transport systems	0.05	769	0.15	2166	0.23	805	0.07	106	0.02	33
<i>Glycolysis</i>												
G	COG0205	6-phosphofructokinase	0.18	2997	0.20	2794	0.10	338	0.12	173	0.20	346
G	COG0837	Glucokinase	0.01	128	0.01	126	0.00	6	0.00	0	0.00	0
G	COG0696	Phosphoglyceromutase	0.05	761	0.04	502	0.03	99	0.07	98	0.05	89
G	COG3635	Predicted phosphoglycerate mutase	0.01	144	0.01	136	0.03	92	0.02	26	0.02	27
<i>Pentose phosphate pathway</i>												
G	COG0021	Transketolase	0.03	481	0.03	481	0.04	124	0.02	30	0.01	16
G	COG0120	Ribose 5-phosphate isomerase	0.04	713	0.06	816	0.02	74	0.01	8	0.03	58
G	COG0362	6-phosphogluconate dehydrogenase	0.03	545	0.01	188	0.01	50	0.00	0	0.01	20
G	COG0363	6-phosphogluconolactonase/Glucosamine-6-phosphate isomerase/deaminase	0.06	905	0.06	827	0.03	118	0.10	138	0.06	111
G	COG0698	Ribose 5-phosphate isomerase RpiB	0.04	681	0.02	314	0.11	372	0.10	143	0.11	200
SIGNAL TRANSDUCTION												
<i>Chemotaxis</i>												
N	COG0840	Methyl-accepting chemotaxis protein	0.13	2187	0.09	1224	0.01	23	0.00	7	0.00	0
N	COG2201	Chemotaxis response regulator containing a CheY-like receiver domain and a methylesterase domain	0.11	1771	0.13	1790	0.01	50	0.00	0	0.01	9
N	COG1871	Chemotaxis protein; stimulates methylation of MCP proteins	0.02	280	0.01	115	0.01	32	0.00	0	0.00	0
N	COG1352	Methylase of chemotaxis methyl-accepting proteins	0.10	1546	0.14	1978	0.02	53	0.01	10	0.03	49
N	COG0643	Chemotaxis protein histidine kinase and related kinases	0.04	721	0.07	921	0.00	13	0.01	13	0.00	0
N	COG0835	Chemotaxis signal transduction protein	0.17	2676	0.15	2187	0.03	95	0.01	20	0.02	29
T	COG0784	FOG: CheY-like receiver	1.25	20303	1.37	19412	0.24	834	0.12	171	0.19	322
N	COG3143	Chemotaxis protein	0.00	8	0.00	0	0.00	11	0.01	13	0.01	11
N	COG1776	Chemotaxis protein CheC	0.03	441	0.02	324	0.02	57	0.00	0	0.00	0
<i>OmpR</i>												
T	COG0642	Signal transduction histidine kinase	2.00	32482	1.78	25210	0.50	1736	0.24	341	0.15	266
T	COG0745	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	0.32	5120	0.25	3506	0.41	1401	0.21	309	0.13	224

Table S4.3: Continued.

Category	COG ID	Subcategory	<i>Microcerotermes</i>		<i>Nasutitermes</i>		<i>Porotermes</i>		<i>Mastotermes_DW</i>		<i>Mastotermes_TV</i>	
			% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*
<i>NtrC</i>												
T	COG3850	Signal transduction histidine kinase	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG3851	Signal transduction histidine kinase	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG4585	Signal transduction histidine kinase	0.01	224	0.01	199	0.01	42	0.03	44	0.03	55
T	COG2197	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	0.21	3461	0.12	1706	0.09	305	0.11	153	0.07	120
<i>NarL</i>												
T	COG3852	Signal transduction histidine kinase	0.00	0	0.00	0	0.00	11	0.01	13	0.01	13
T	COG0642	Signal transduction histidine kinase	2.00	32482	1.78	25210	0.50	1736	0.24	341	0.15	266
S	COG3304	Predicted membrane protein	0.00	24	0.00	42	0.00	15	0.02	24	0.02	29
<i>CitB</i>												
T	COG3290	Signal transduction histidine kinase regulating citrate/malate metabolism	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
K	COG4565	Response regulator of citrate/malate metabolism	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
<i>LytT/AgrA</i>												
T	COG2972	Predicted signal transduction protein with a C-terminal ATPase domain	0.06	897	0.04	576	0.00	6	0.00	0	0.00	0
T	COG3275	Putative regulator of cell autolysis	0.00	0	0.00	0	0.00	8	0.00	0	0.00	0
K	COG3279	Response regulator of the LytR/AlgR family	0.07	1138	0.03	429	0.03	92	0.09	124	0.04	73
<i>Other kinases, phosphatases</i>												
T	COG2205	Osmosensitive K ⁺ channel histidine kinase	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG3920	Signal transduction histidine kinase	0.00	8	0.00	10	0.00	4	0.01	10	0.00	0
T	COG4191	Signal transduction histidine kinase regulating C4-dicarboxylate transport system	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG4192	Signal transduction histidine kinase regulating phosphoglycerate transport system	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG4251	Bacteriophytochrome (light-regulated signal transduction histidine kinase)	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG4564	Signal transduction histidine kinase	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG5000	Signal transduction histidine kinase involved in nitrogen fixation and metabolism regulation	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG5002	Signal transduction histidine kinase	0.00	64	0.00	21	0.01	27	0.00	0	0.00	0
T	COG0478	RIO-like serine/threonine protein kinase fused to N-terminal HTH domain	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
R	COG0515	Serine/threonine protein kinase	0.08	1218	0.06	858	0.07	246	0.07	105	0.08	142
T	COG0317	Guanosine polyphosphate pyrophosphohydrolases/synthetases	0.06	1042	0.05	764	0.07	231	0.08	121	0.05	84
T	COG0394	Protein-tyrosine-phosphatase	0.01	144	0.01	73	0.01	34	0.00	0	0.00	4
T	COG0467	RecA-superfamily ATPases implicated in signal transduction	0.00	56	0.00	63	0.00	0	0.00	0	0.00	4
T	COG0631	Serine/threonine protein phosphatase	0.04	625	0.03	429	0.05	170	0.07	103	0.05	80
G	COG1762	Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)	0.21	3421	0.25	3569	0.06	210	0.08	110	0.14	246

Table S4.3: Continued.

Category	COG ID	Subcategory	Microcerotermes		Nasutitermes		Porotermes		Mastotermes_DW		Mastotermes_TV	
			% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*
T	COG2062	Phosphohistidine phosphatase SixA	0.00	8	0.00	0	0.02	59	0.00	7	0.00	0
T	COG2114	Adenylate cyclase	0.36	5865	0.31	4416	0.04	145	0.01	13	0.06	104
T	COG2365	Protein tyrosine/serine phosphatase	0.01	216	0.01	105	0.01	21	0.00	4	0.00	0
T	COG2453	Predicted protein-tyrosine phosphatase	0.00	8	0.00	0	0.00	0	0.01	18	0.00	0
<i>Other regulators, domains</i>												
T	COG1551	Carbon storage regulator (could also regulate swarming and quorum sensing)	0.06	1002	0.09	1224	0.02	53	0.00	4	0.02	33
T	COG1639	Predicted signal transduction protein	0.07	1178	0.07	984	0.01	29	0.00	0	0.01	20
T	COG1702	Phosphate starvation-inducible protein PhoH	0.05	801	0.04	576	0.03	116	0.05	66	0.01	24
T	COG1956	GAF domain-containing protein	0.00	0	0.00	52	0.02	65	0.03	38	0.02	31
K	COG1974	SOS-response transcriptional repressors (RecA-mediated autopeptidases)	0.10	1570	0.11	1538	0.04	124	0.00	5	0.11	200
T	COG2172	Anti-sigma regulatory factor (Ser/Thr protein kinase)	0.08	1362	0.09	1203	0.03	95	0.01	14	0.01	11
T	COG1716	FOG: FHA domain	0.00	24	0.01	178	0.01	40	0.06	88	0.01	20
T	COG2198	FOG: HPt domain	0.01	112	0.02	251	0.00	2	0.00	0	0.00	0
T	COG2199	FOG: GGDEF domain	0.35	5673	0.55	7796	0.09	307	0.09	133	0.04	69
T	COG2200	FOG: EAL domain	0.05	745	0.11	1496	0.01	50	0.00	0	0.02	27
T	COG2203	FOG: GAF domain	0.02	377	0.01	167	0.01	42	0.00	0	0.01	9
T	COG2206	HD-GYP domain	0.30	4863	0.29	4144	0.04	153	0.01	13	0.04	69
T	COG2770	FOG: HAMP domain	0.00	8	0.00	10	0.00	0	0.00	0	0.00	0
T	COG2905	Predicted signal-transduction protein containing cAMP-binding and CBS domains	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG3434	Predicted signal transduction protein containing EAL and modified HD-GYP domains	0.01	104	0.01	105	0.00	0	0.00	0	0.00	0
K	COG3437	Response regulator containing a CheY-like receiver domain and an HD-GYP domain	0.41	6690	0.36	5107	0.07	254	0.01	13	0.02	33
T	COG3456	Uncharacterized conserved protein	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG3480	Predicted secreted protein containing a PDZ domain	0.00	0	0.00	31	0.01	21	0.00	0	0.00	0
K	COG3604	Transcriptional regulator containing GAF	0.04	641	0.01	73	0.02	74	0.03	44	0.02	42
T	COG3605	Signal transduction protein containing GAF and PtsI domains	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG3629	DNA-binding transcriptional activator of the SARP family	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG3706	Response regulator containing a CheY-like receiver domain and a GGDEF domain	0.14	2284	0.04	534	0.01	50	0.00	7	0.00	0
T	COG3707	Response regulator with putative antiterminator output domain	0.01	120	0.00	10	0.00	11	0.04	51	0.07	113
T	COG3830	ACT domain-containing protein	0.07	1066	0.06	816	0.03	120	0.03	45	0.03	53
T	COG3887	Predicted signaling protein consisting of a modified GGDEF domain and a DHH domain	0.00	8	0.00	42	0.01	38	0.00	0	0.00	0
T	COG3947	Response regulator containing CheY-like receiver and SARP domains	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG4566	Response regulator	0.00	0	0.00	0	0.00	11	0.02	24	0.02	27
T	COG4567	Response regulator consisting of a CheY-like receiver domain and a Fis-type HTH domain	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG4725	Transcriptional activator	0.01	112	0.00	42	0.01	32	0.00	0	0.00	0

Table S4.3: Continued.

Category	COG ID	Subcategory	Microcerotermes		Nasutitermes		Porotermes		Mastotermes_DW		Mastotermes_TV	
			% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*
T	COG4753	Response regulator containing CheY-like receiver domain and AraC-type DNA-binding domain	0.01	144	0.00	10	0.00	6	0.01	11	0.00	0
K	COG4978	Transcriptional regulator	0.00	0	0.00	10	0.00	6	0.00	6	0.00	0
<i>Sensors, domains</i>												
T	COG1966	Carbon starvation protein	0.05	841	0.07	931	0.05	179	0.01	19	0.02	42
T	COG2202	FOG: PAS/PAC domain	0.01	112	0.01	94	0.00	8	0.00	6	0.01	13
T	COG3292	Predicted periplasmic ligand-binding sensor domain	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG3300	MHYT domain (predicted integral membrane sensor domain)	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG3322	Predicted periplasmic ligand-binding sensor domain	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG3447	Predicted integral membrane sensor domain	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG3448	CBS-domain-containing membrane protein	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG3452	Predicted periplasmic ligand-binding sensor domain	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG3614	Predicted periplasmic ligand-binding sensor domain	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
K	COG4219	Antirepressor regulating drug resistance	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG4250	Predicted sensor protein/domain	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG4252	Predicted transmembrane sensor domain	0.00	0	0.00	21	0.00	4	0.00	0	0.00	0
T	COG4936	Predicted sensor domain	0.00	64	0.00	0	0.00	2	0.01	17	0.00	0
T	COG4943	Predicted signal transduction protein containing sensor and EAL domains	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG5001	Predicted signal transduction protein containing a membrane domain	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
T	COG5278	Predicted periplasmic ligand-binding sensor domain	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
CELL MOTILITY												
<i>Flagellum structure and biogenesis</i>												
N	COG1334	Uncharacterized flagellar protein FlaG	0.01	200	0.01	199	0.00	2	0.00	0	0.00	0
N	COG1344	Flagellin and related hook-associated proteins	0.21	3405	0.23	3234	0.02	67	0.01	14	0.06	111
N	COG1419	Flagellar GTP-binding protein	0.06	921	0.06	827	0.01	27	0.00	0	0.02	38
N	COG1536	Flagellar motor switch protein	0.10	1683	0.11	1538	0.01	46	0.01	13	0.03	55
N	COG1580	Flagellar basal body-associated protein	0.05	737	0.08	1162	0.00	13	0.01	10	0.03	53
N	COG1749	Flagellar hook protein FlgE	0.05	809	0.04	534	0.01	36	0.01	16	0.03	55
N	COG1815	Flagellar basal body protein	0.09	1466	0.09	1287	0.02	59	0.01	10	0.01	13
N	COG1843	Flagellar hook capping protein	0.07	1098	0.05	680	0.01	32	0.00	0	0.04	62
N	COG1868	Flagellar motor switch protein	0.10	1562	0.07	1015	0.01	50	0.00	4	0.04	64
N	COG1157	Flagellar biosynthesis/type III secretory pathway ATPase	0.05	889	0.07	1046	0.01	34	0.02	28	0.03	44
N	COG1317	Flagellar biosynthesis/type III secretory pathway protein	0.05	881	0.05	774	0.01	21	0.01	13	0.02	38
N	COG1766	Flagellar biosynthesis/type III secretory pathway lipoprotein	0.06	897	0.07	1026	0.00	11	0.01	13	0.01	20
N	COG1886	Flagellar motor switch/type III secretory pathway protein	0.06	1002	0.07	984	0.01	29	0.00	6	0.04	62
N	COG1987	Flagellar biosynthesis pathway	0.08	1290	0.08	1109	0.02	55	0.00	7	0.03	44

Table S4.3: Continued.

Category	COG ID	Subcategory	Microcerotermes		Nasutitermes		Porotermes		Mastotermes_DW		Mastotermes_TV	
			% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*	% of COG genes^	Gene copy*
<i>Pilus</i>												
N	COG3063	Tfp pilus assembly protein PilF	0.00	0	0.00	0	0.00	6	0.00	5	0.00	0
N	COG4972	Tfp pilus assembly protein	0.00	8	0.00	0	0.00	6	0.00	7	0.00	0
N	COG3166	Tfp pilus assembly protein PilN	0.00	0	0.00	0	0.00	6	0.00	0	0.01	9
N	COG3188	P pilus assembly protein	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
N	COG2165	Type II secretory pathway	0.02	304	0.02	324	0.03	90	0.03	37	0.01	16
N	COG2804	Type II secretory pathway	0.06	897	0.03	387	0.04	147	0.02	30	0.00	4
N	COG2805	Tfp pilus assembly protein	0.06	929	0.02	324	0.05	170	0.03	48	0.01	13
DEFENCE												
V	COG4096	Type I site-specific restriction-modification system	0.02	377	0.02	324	0.01	19	0.00	7	0.00	0
V	COG4452	Inner membrane protein involved in colicin E2 resistance	0.03	553	0.02	251	0.02	71	0.01	14	0.00	0
V	COG0841	Cation/multidrug efflux pump	0.10	1667	0.11	1538	0.11	385	0.13	190	0.14	251
V	COG0286	Type I restriction-modification system methyltransferase subunit	0.10	1659	0.13	1831	0.07	250	0.03	42	0.01	13
V	COG4823	Abortive infection bacteriophage resistance protein	0.01	96	0.02	230	0.00	4	0.01	10	0.00	4
V	COG2720	Uncharacterized vancomycin resistance protein	0.00	56	0.00	10	0.01	19	0.00	0	0.00	0
V	COG1619	Uncharacterized proteins	0.01	112	0.01	94	0.00	8	0.01	12	0.00	0
V	COG0732	Restriction endonuclease S subunits	0.02	401	0.02	324	0.02	78	0.01	10	0.00	0
V	COG1403	Restriction endonuclease	0.02	264	0.01	147	0.01	42	0.02	31	0.03	53
V	COG1136	ABC-type antimicrobial peptide transport system	0.38	6194	0.33	4688	0.40	1378	0.27	393	0.40	690
V	COG1131	ABC-type multidrug transport system	0.22	3541	0.17	2438	0.33	1156	0.11	158	0.22	386
V	COG2274	ABC-type bacteriocin/lantibiotic exporters	0.04	633	0.01	136	0.01	50	0.04	59	0.01	11

^ Percentage of total COG assigned genes

* Gene count is the number of genes weighted by their individual read depth

Table S4.4: Major functions represented between the metagenomes based on Kyoto Encyclopedia of Genes and Genomes (KEGG).

Category	KO ID	<i>Microcerotermes</i> Gene copy*	<i>Nasutitermes</i> Gene copy*	<i>Porotermes</i> Gene copy*	<i>Mastotermes_DW</i> Gene copy*	<i>Mastotermes_TV</i> Gene copy*
ENERGY METABOLISM						
<i>Nitrogen fixation</i>						
nitrogen fixation protein NifB	K02585	52.6	58.6	28.6	72.1	68.9
nitrogenase molybdenum-iron protein alpha chain nifD	K02586	170.8	468.9	85.7	113.2	160.9
nitrogenase molybdenum-cofactor synthesis protein nifE	K02587	131.4	136.8	17.1	61.8	68.9
nitrogenase iron protein nifH	K02588	197.1	254.0	51.4	72.1	103.4
nitrogen regulatory protein PII 1 nifDH1	K02589	92.0	39.1	17.1	41.2	46.0
nitrogen regulatory protein PII 1 nifDH1	K02590	92.0	19.5	17.1	41.2	68.9
nitrogenase molybdenum-iron protein beta chain nifK	K02591	249.7	508.0	74.3	164.7	126.4
<i>Homoacetogenesis (Wood-Ljungdahl pathway)</i>						
carbon-monoxide dehydrogenase catalytic subunit [EC:1.2.99.2 1.2.7.4]	K00198	52.6	117.2	5.7	10.3	46.0
formate--tetrahydrofolate ligase [EC:6.3.4.3]	K01938	39.4	214.9	34.3	61.8	103.4
methylenetetrahydrofolate dehydrogenase (NADP+) / methenyltetrahydrofolate cyclohydrolase [EC:1.5.1.5 3.5.4.9]	K01491	105.1	175.8	68.5	113.2	229.8
methylenetetrahydrofolate reductase (NADPH) [EC:1.5.1.20]	K00297	118.3	214.9	40.0	41.2	103.4
5-methyltetrahydrofolate corrinoid [EC:2.1.1.258]	K15023	0.0	0.0	11.4	30.9	11.5
acetyl-CoA synthase [EC:2.3.1.169]	K14138	13.1	39.1	11.4	61.8	11.5
acetyl-CoA decarbonylase/synthase complex subunit gamma [EC:2.1.1.245]	K00197	52.6	0.0	17.1	72.1	23.0
acetyl-CoA decarbonylase/synthase complex subunit delta [EC:2.1.1.245]	K00194	26.3	19.5	5.7	41.2	34.5
<i>Pyruvate metabolism</i>						
phosphoenolpyruvate carboxylase [EC 4.1.1.31]	K01595	0.0	0.0	5.7	0.0	11.5
oxaloacetate decarboxylase, alpha subunit [EC 4.1.1.3]	K01571	52.6	214.9	34.3	20.6	80.4
oxaloacetate decarboxylase, beta subunit [EC 4.1.1.3]	K01572	92.0	351.7	34.3	113.2	149.4
pyruvate carboxylase subunit A [EC 6.4.1.1]	K01959	13.1	19.5	0.0	0.0	0.0
pyruvate carboxylase subunit B [EC 6.4.1.1]	K01960	52.6	19.5	22.8	72.1	23.0
malate dehydrogenase [EC 1.1.1.37]	K00024	52.6	19.5	28.6	30.9	80.4
malate dehydrogenase (oxaloacetate-decarboxylating)(NADP+)[EC 1.1.1.40]	K00029	13.1	19.5	17.1	41.2	68.9
fumarate hydratase, class I [EC 4.2.1.2]	K01676	0.0	39.1	17.1	51.5	80.4
fumarate hydratase subunit alpha [EC 4.2.1.2]	K01677	0.0	156.3	28.6	41.2	34.5
fumarate hydratase subunit beta [EC 4.2.1.2]	K01678	26.3	97.7	17.1	41.2	34.5
formate C-acetyltransferase [EC 2.3.1.54]	K00656	39.4	39.1	62.8	205.9	160.9
pyruvate formate lyase activating enzyme [EC 1.97.1.4]	K04069	210.3	254.0	62.8	175.0	57.5
putative pyruvate-flavodoxin oxidoreductase [EC 1.2.7.-]	K03737	131.4	234.5	80.0	164.7	229.8
acetyl-CoA synthetase [EC 6.2.1.1]	K01895	262.8	175.8	22.8	102.9	80.4
phosphate acetyltransferase [EC 2.3.1.8]	K00625	78.8	78.2	51.4	102.9	149.4
acetate kinase [EC 2.7.2.1]	K00925	144.6	117.2	74.3	175.0	114.9
acetaldehyde dehydrogenase / alcohol dehydrogenase [EC 1.2.1.10 1.1.1.1]	K04072	13.1	214.9	34.3	20.6	23.0

Table S4.4: Continued.

Category	KO ID	<i>Microcerotermes</i> Gene copy*	<i>Nasutitermes</i> Gene copy*	<i>Porotermes</i> Gene copy*	<i>Mastotermes_DW</i> Gene copy*	<i>Mastotermes_TV</i> Gene copy*
ENERGY METABOLISM						
<i>Nitrogen fixation</i>						
nitrogen fixation protein NifB	K02585	52.6	58.6	28.6	72.1	68.9
nitrogenase molybdenum-iron protein alpha chain nifD	K02586	170.8	468.9	85.7	113.2	160.9
nitrogenase molybdenum-cofactor synthesis protein nifE	K02587	131.4	136.8	17.1	61.8	68.9
nitrogenase iron protein nifH	K02588	197.1	254.0	51.4	72.1	103.4
nitrogen regulatory protein PII 1 nifDH1	K02589	92.0	39.1	17.1	41.2	46.0
nitrogen regulatory protein PII 1 nifDH1	K02590	92.0	19.5	17.1	41.2	68.9
nitrogenase molybdenum-iron protein beta chain nifK	K02591	249.7	508.0	74.3	164.7	126.4
<i>Homoacteogenesis (Wood-Ljungdahl pathway)</i>						
carbon-monoxide dehydrogenase catalytic subunit [EC:1.2.99.2 1.2.7.4]	K00198	52.6	117.2	5.7	10.3	46.0
formate--tetrahydrofolate ligase [EC:6.3.4.3]	K01938	39.4	214.9	34.3	61.8	103.4
methylenetetrahydrofolate dehydrogenase (NADP+) / methenyltetrahydrofolate cyclohydrolase [EC:1.5.1.5 3.5.4.9]	K01491	105.1	175.8	68.5	113.2	229.8
methylenetetrahydrofolate reductase (NADPH) [EC:1.5.1.20]	K00297	118.3	214.9	40.0	41.2	103.4
5-methyltetrahydrofolate corrinoid [EC:2.1.1.258]	K15023	0.0	0.0	11.4	30.9	11.5
acetyl-CoA synthase [EC:2.3.1.169]	K14138	13.1	39.1	11.4	61.8	11.5
acetyl-CoA decarboxylase/synthase complex subunit gamma [EC:2.1.1.245]	K00197	52.6	0.0	17.1	72.1	23.0
acetyl-CoA decarboxylase/synthase complex subunit delta [EC:2.1.1.245]	K00194	26.3	19.5	5.7	41.2	34.5
formate dehydrogenase major subunit [EC:1.2.1.2]	K00123	0.0	0.0	45.7	82.4	34.5
<i>Pyruvate metabolism</i>						
phosphoenolpyruvate carboxylase [EC 4.1.1.31]	K01595	0.0	0.0	5.7	0.0	11.5
oxaloacetate decarboxylase, alpha subunit [EC 4.1.1.3]	K01571	52.6	214.9	34.3	20.6	80.4
oxaloacetate decarboxylase, beta subunit [EC 4.1.1.3]	K01572	92.0	351.7	34.3	113.2	149.4
pyruvate carboxylase subunit A [EC 6.4.1.1]	K01959	13.1	19.5	0.0	0.0	0.0
pyruvate carboxylase subunit B [EC 6.4.1.1]	K01960	52.6	19.5	22.8	72.1	23.0
malate dehydrogenase [EC 1.1.1.37]	K00024	52.6	19.5	28.6	30.9	80.4
malate dehydrogenase (oxaloacetate-decarboxylating)(NADP+)[EC 1.1.1.40]	K00029	13.1	19.5	17.1	41.2	68.9
fumarate hydratase, class I [EC 4.2.1.2]	K01676	0.0	39.1	17.1	51.5	80.4
fumarate hydratase subunit alpha [EC 4.2.1.2]	K01677	0.0	156.3	28.6	41.2	34.5
fumarate hydratase subunit beta [EC 4.2.1.2]	K01678	26.3	97.7	17.1	41.2	34.5
formate C-acetyltransferase [EC 2.3.1.54]	K00656	39.4	39.1	62.8	205.9	160.9
pyruvate formate lyase activating enzyme [EC 1.97.1.4]	K04069	210.3	254.0	62.8	175.0	57.5
putative pyruvate-flavodoxin oxidoreductase [EC 1.2.7.-]	K03737	131.4	234.5	80.0	164.7	229.8
acetyl-CoA synthetase [EC 6.2.1.1]	K01895	262.8	175.8	22.8	102.9	80.4
phosphate acetyltransferase [EC 2.3.1.8]	K00625	78.8	78.2	51.4	102.9	149.4
acetate kinase [EC 2.7.2.1]	K00925	144.6	117.2	74.3	175.0	114.9
acetaldehyde dehydrogenase / alcohol dehydrogenase [EC 1.2.1.10 1.1.1.1]	K04072	13.1	214.9	34.3	20.6	23.0

Table S4.4: Continued.

Category	KO ID	<i>Microcerotermes</i> Gene copy*	<i>Nasutitermes</i> Gene copy*	<i>Porotermes</i> Gene copy*	<i>Mastotermes_DW</i> Gene copy*	<i>Mastotermes_TV</i> Gene copy*
CARBHYDRATE METABOLISM						
<i>Glycolysis</i>						
hexokinase [EC:2.7.1.1]	K00844	92.0	117.2	17.1	10.3	114.9
glucose-6-phosphate isomerase [EC:5.3.1.9]	K01810	65.7	214.9	62.8	72.1	137.9
6-phosphofructokinase 1 [EC:2.7.1.11]	K00850	223.4	312.6	57.1	144.1	218.3
fructose-bisphosphate aldolase, class I [EC:4.1.2.13]	K11645	0.0	0.0	11.4	0.0	0.0
fructose-bisphosphate aldolase, class II [EC:4.1.2.13]	K01624	92.0	195.4	62.8	164.7	149.4
triosephosphate isomerase (TIM) [EC:5.3.1.1]	K01803	249.7	449.4	62.8	82.4	160.9
glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12]	K00134	39.4	97.7	57.1	113.2	160.9
phosphoglycerate kinase [EC:2.7.2.3]	K00927	39.4	78.2	62.8	123.5	149.4
2,3-bisphosphoglycerate-dependent phosphoglycerate mutase [EC:5.4.2.11]	K01834	13.1	58.6	22.8	61.8	23.0
2,3-bisphosphoglycerate-independent phosphoglycerate mutase [EC:5.4.2.12]	K15633	157.7	254.0	28.6	123.5	160.9
probable phosphoglycerate mutase [EC:5.4.2.12]	K15634	0.0	58.6	28.6	20.6	0.0
2,3-bisphosphoglycerate-independent phosphoglycerate mutase [EC:5.4.2.12]	K15635	13.1	39.1	22.8	72.1	68.9
enolase [EC:4.2.1.11]	K01689	78.8	97.7	51.4	185.3	103.4
pyruvate kinase [EC:2.7.1.40]	K00873	39.4	117.2	28.6	82.4	103.4
phosphoglucomutase [EC:5.4.2.2]	K01835	52.6	254.0	28.6	123.5	206.8
<i>Galactose metabolism</i>						
alpha-glucosidase [EC 3.2.1.20]	K01187	13.1	58.6	74.3	92.7	137.9
beta-galactosidase [EC 3.2.1.23]	K01190	105.1	175.8	114.2	226.5	459.7
UDP-glucose 4-epimerase [EC 5.1.3.2]	K01784	170.8	312.6	62.8	133.8	126.4
UDP-galactopyranose mutase [EC 5.4.99.9]	K01854	13.1	58.6	17.1	20.6	57.5
alpha-galactosidase [EC 3.2.1.22]	K07407	39.4	19.5	45.7	102.9	80.4
UDP-glucose 4-epimerase [EC 5.1.3.2]	K17716	0.0	0.0	5.7	0.0	0.0
<i>Chitin metabolism</i>						
chitinase [EC 3.2.1.14]	K01183	52.6	97.7	5.7	20.6	0.0
beta-N-acetylhexosaminidase [EC 3.2.1.52]	K01207	118.3	312.6	28.6	10.3	46.0
N-acetylglucosamine-1-P-mutase (Phosphomannomutase) [EC 5.4.2.8]	K01840	39.4	78.2	11.4	82.4	91.9
glucokinase [EC 2.7.1.2]	K00845	26.3	58.6	34.3	175.0	68.9
glucosamine-6-phosphate deaminase [EC 3.5.99.6]	K02564	39.4	214.9	34.3	133.8	172.4
CELL MOTILITY						
<i>Flagellar assembly</i>						
flagella basal body P-ring formation protein FlgA	K02386	13.1	78.2	0.0	10.3	11.5
flagellar basal-body rod protein FlgB	K02387	131.4	214.9	0.0	0.0	23.0
flagellar basal-body rod protein FlgC	K02388	144.6	175.8	5.7	20.6	11.5

Table S4.4: Continued.

Category	KO ID	<i>Microcerotermes</i> Gene copy*	<i>Nasutitermes</i> Gene copy*	<i>Porotermes</i> Gene copy*	<i>Mastotermes_DW</i> Gene copy*	<i>Mastotermes_TV</i> Gene copy*
CELL MOTILITY						
flagellar basal-body rod modification protein FlgD	K02389	39.4	136.8	5.7	0.0	23.0
flagellar hook protein FlgE	K02390	131.4	214.9	11.4	10.3	34.5
flagellar basal-body rod protein FlgF	K02391	0.0	0.0	0.0	0.0	0.0
flagellar basal-body rod protein FlgG	K02392	354.8	468.9	11.4	30.9	34.5
flagellar L-ring protein precursor FlgH	K02393	0.0	39.1	0.0	10.3	23.0
flagellar P-ring protein precursor FlgI	K02394	0.0	39.1	5.7	0.0	11.5
flagellar hook-associated protein 1 FlgK	K02396	197.1	254.0	11.4	10.3	34.5
flagellar hook-associated protein 3 FlgL	K02397	131.4	293.1	0.0	0.0	0.0
negative regulator of flagellin synthesis FlgM	K02398	78.8	58.6	5.7	0.0	11.5
flagellar biosynthesis protein FlhA	K02400	276.0	449.4	17.1	41.2	34.5
flagellar biosynthetic protein FlhB	K02401	92.0	254.0	22.8	10.3	23.0
flagellar transcriptional activator FlhC	K02402	0.0	0.0	0.0	0.0	0.0
flagellar transcriptional activator FlhD	K02403	0.0	0.0	0.0	0.0	0.0
flagellin, FliC	K02406	262.8	605.7	45.7	10.3	46.0
flagellar hook-associated protein 2, FliD	K02407	78.8	214.9	11.4	10.3	23.0
flagellar hook-basal body complex protein FliE	K02408	131.4	175.8	5.7	10.3	0.0
flagellar M-ring protein FliF	K02409	157.7	254.0	11.4	41.2	23.0
flagellar motor switch protein FliG	K02410	328.5	605.7	40.0	10.3	23.0
flagellar assembly protein FliH	K02411	118.3	195.4	5.7	10.3	11.5
flagellum-specific ATP synthase [EC:3.6.3.14]	K02412	157.7	156.3	17.1	10.3	11.5
flagellar FliJ protein	K02413	105.1	136.8	11.4	10.3	23.0
flagellar hook-length control protein FliK	K02414	0.0	19.5	0.0	0.0	0.0
flagellar motor switch protein FliM	K02416	249.7	214.9	0.0	0.0	11.5
flagellar motor switch protein FliN/FliY	K02417	170.8	293.1	0.0	0.0	11.5
flagellar protein FliO/FliZ	K02418	92.0	117.2	0.0	0.0	11.5
flagellar biosynthetic protein FliP	K02419	118.3	156.3	11.4	0.0	11.5
flagellar biosynthetic protein FliQ	K02420	78.8	156.3	11.4	0.0	23.0
flagellar biosynthetic protein FliR	K02421	92.0	175.8	17.1	10.3	0.0
flagellar protein FliS	K02422	144.6	175.8	0.0	0.0	11.5
flagellar protein FliT	K02423	0.0	0.0	5.7	0.0	0.0
flagellar motor A MotA	K02556	131.4	351.7	5.7	20.6	34.5
flagellar motor B MotB	K02557	105.1	234.5	17.1	20.6	34.5
flagella synthesis protein FlgN	K20399	0.0	0.0	0.0	0.0	0.0
<i>Bacterial chemotaxis</i>						
methyl-accepting chemotaxis protein	K03406	2707.1	6486.7	120.0	92.7	114.9
chemotaxis protein CheD	K03411	13.1	19.5	5.7	0.0	0.0
chemotaxis protein methyltransferase CheR	K00575	170.8	390.8	5.7	0.0	57.5

Table S4.4: Continued.

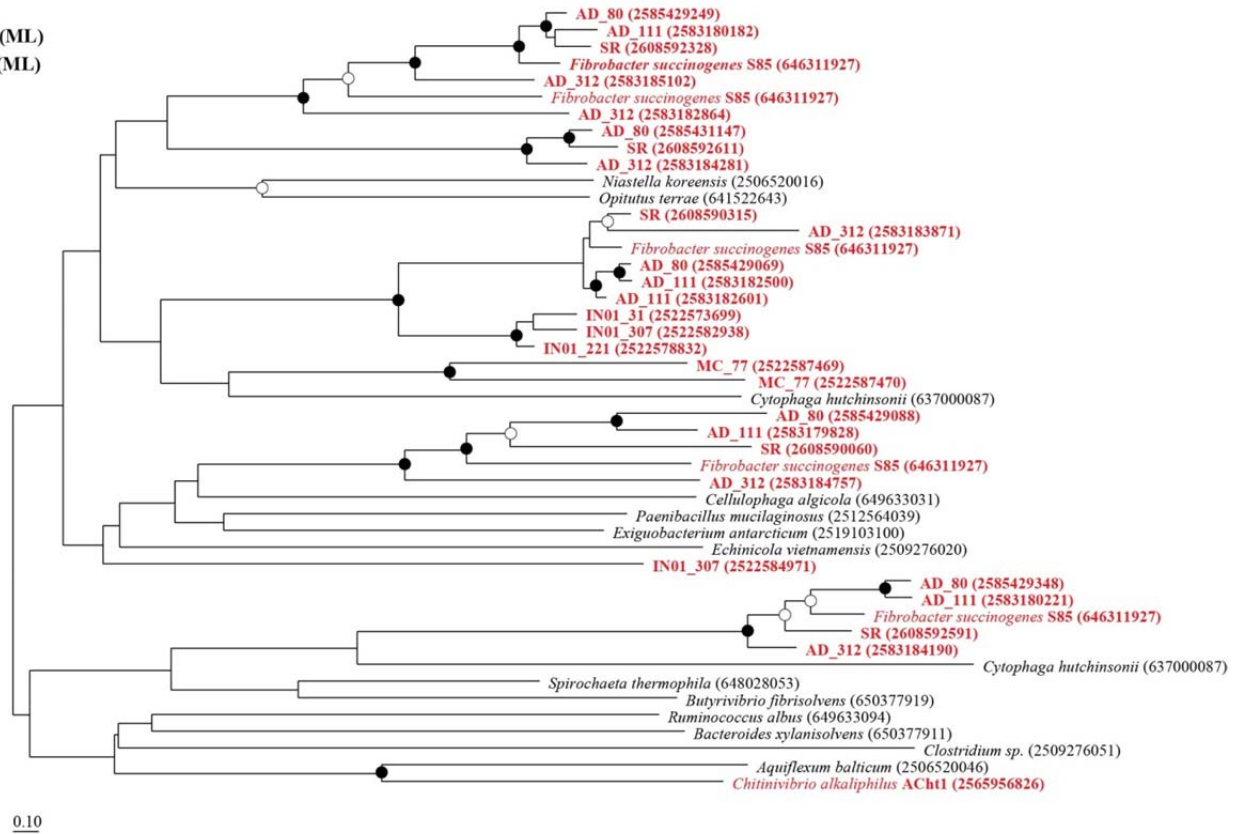
Category	KO ID	<i>Microcerotermes</i> Gene copy*	<i>Nasutitermes</i> Gene copy*	<i>Porotermes</i> Gene copy*	<i>Mastotermes_DW</i> Gene copy*	<i>Mastotermes_TV</i> Gene copy*
CELL MOTILITY						
two-component system, chemotaxis family, response regulator CheB	K03412	170.8	429.8	0.0	0.0	46.0
two-component system, chemotaxis family, sensor kinase CheA	K03407	315.4	508.0	11.4	10.3	23.0
purine-binding chemotaxis protein CheW	K03408	223.4	429.8	17.1	0.0	34.5
two-component system, chemotaxis family, response regulator CheV	K03415	13.1	19.5	0.0	0.0	23.0
two-component system, chemotaxis family, response regulator CheY	K03413	223.4	273.5	17.1	10.3	23.0
chemotaxis protein CheC	K03410	13.1	19.5	0.0	0.0	0.0
chemotaxis protein CheZ	K03414	0.0	0.0	5.7	10.3	0.0
chemotaxis protein CheX	K03409	223.4	254.0	34.3	10.3	23.0

* Gene count is the number of genes weighted by their individual read depth

Appendix D: Supplementary figures and tables for Chapter 5

A. CBM4

- ≥ 75% (ML)
- 50-74% (ML)



B. CBM48

- ≥ 75% (ML)
- 50-74% (ML)

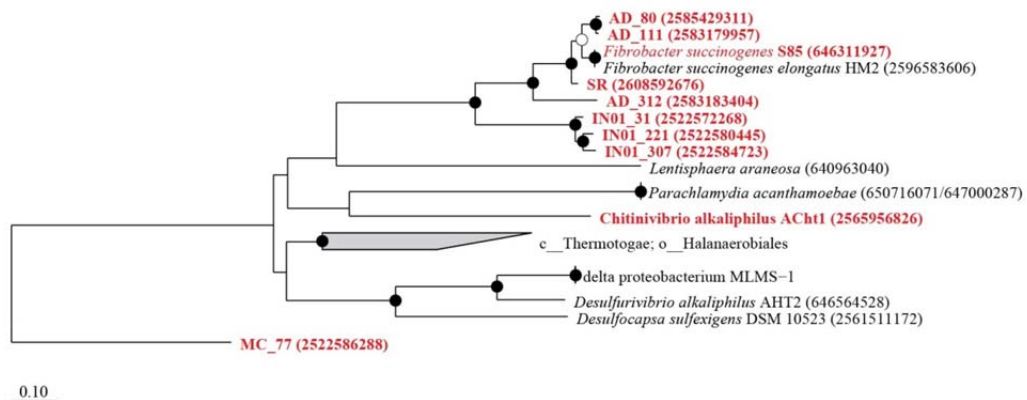


Figure S5.1: A maximum likelihood phylogenetic analysis of two CBM families; A) CBM4 and B) CBM46 > 722aa. Carbohydrate binding module (CBM) protein sequences were aligned using MAFFT v7.221 and phylogenetic trees were constructed from 2,256 finished genomes from the IMG database (Markowitz et al., 2009) using Fasttree v2.1.7. The trees are unrooted and only the closest neighbors of the *Fibrobacteres* (in red) are shown with corresponding IMG IDs in brackets. Bootstrap support for interior nodes is indicated by dots according to the legend at the top left of the figure.

Fibro-slime domain protein

- ≥ 75% (ML)
- 50-74% (ML)

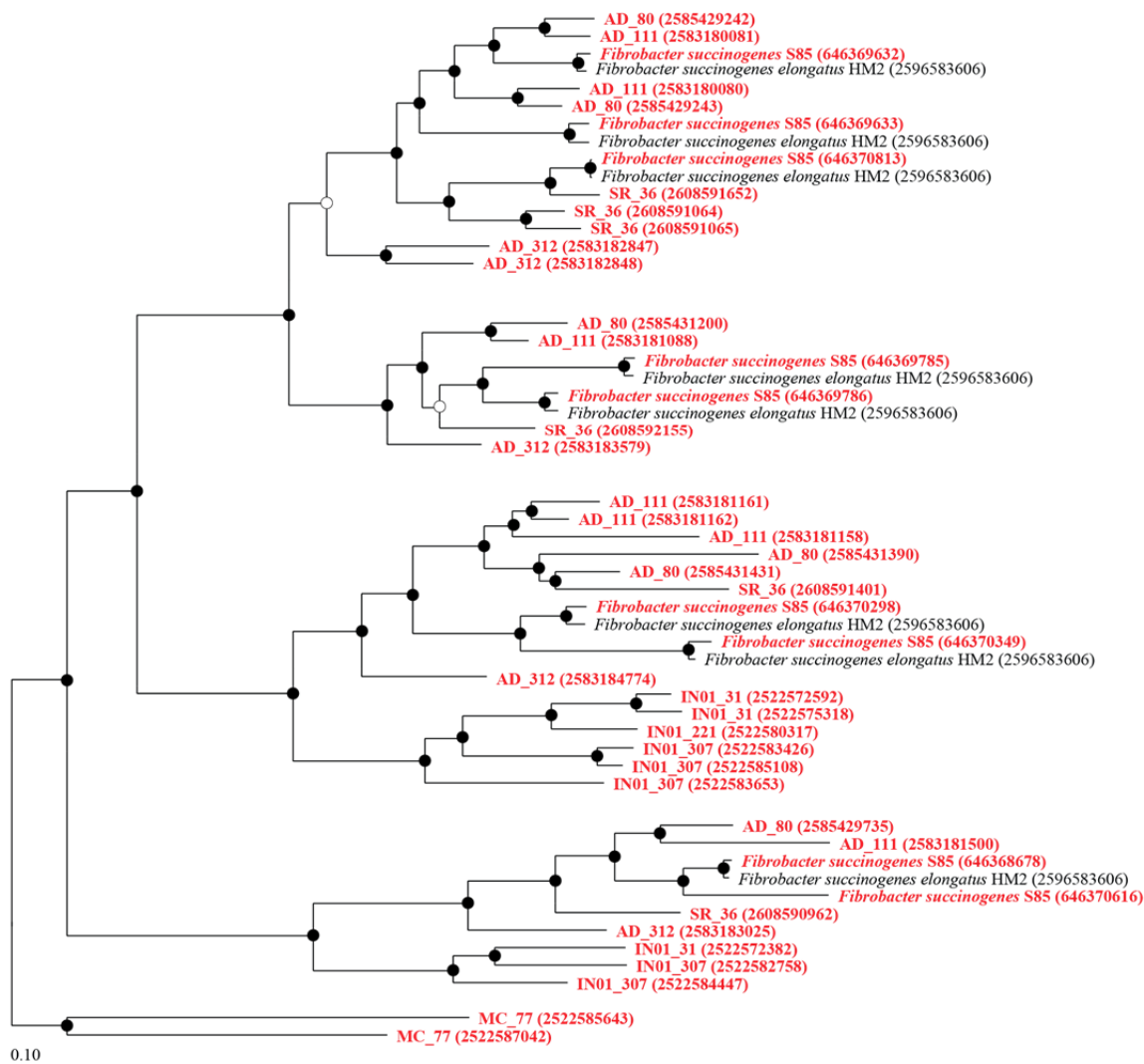
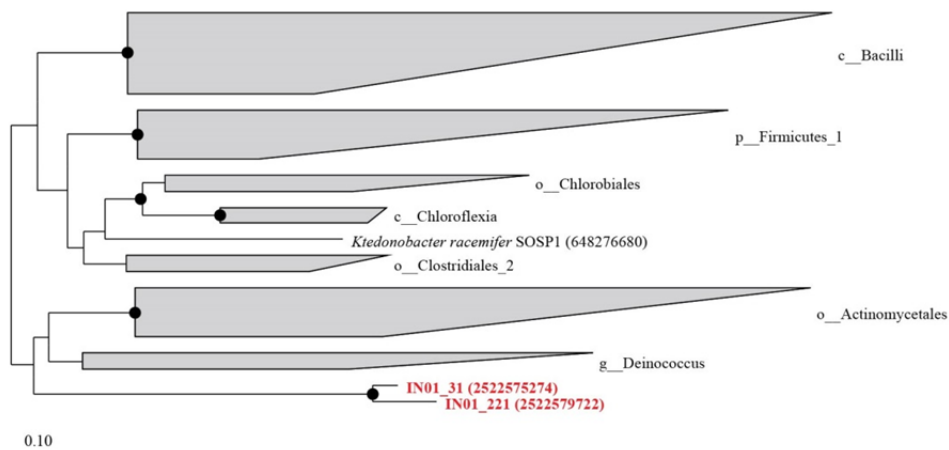


Figure S5.2: A maximum likelihood phylogenetic analysis of fibro-slime domain proteins >600aa. Fibro-slime domain protein sequences were aligned using MAFFT v7.221 and phylogenetic trees were constructed from 2,256 finished genomes from the IMG database (Markowitz et al., 2009) using Fasttree v2.1.7. The tree is rooted with MC_77 and corresponding IMG IDs shown in parentheses. Bootstrap support for interior nodes is indicated by dots according to the legend at the top left of the figure.

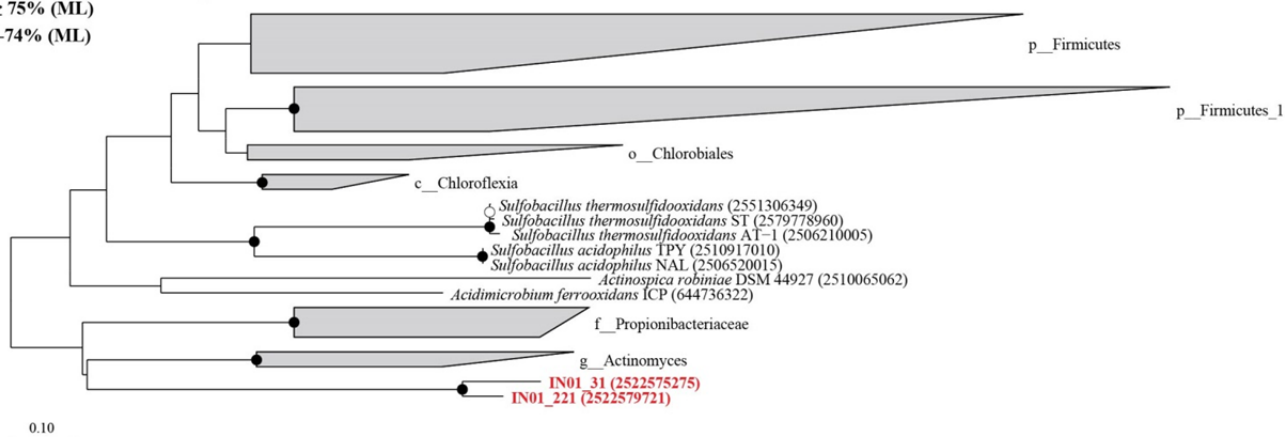
A. Cytochrome bd-I ubiquinol oxidase subunit 1

- ≥ 75% (ML)
- 50-74% (ML)



B. Cytochrome bd-I ubiquinol oxidase subunit 2

- ≥ 75% (ML)
- 50-74% (ML)



C. ABC-type transport system involved in cytochrome bd biosynthesis, fused ATPase and permease components

- ≥ 75% (ML)
- 50-74% (ML)

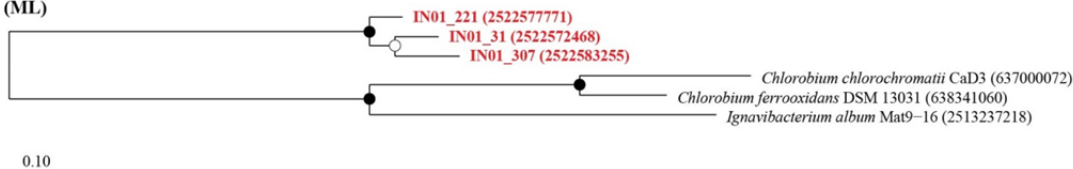
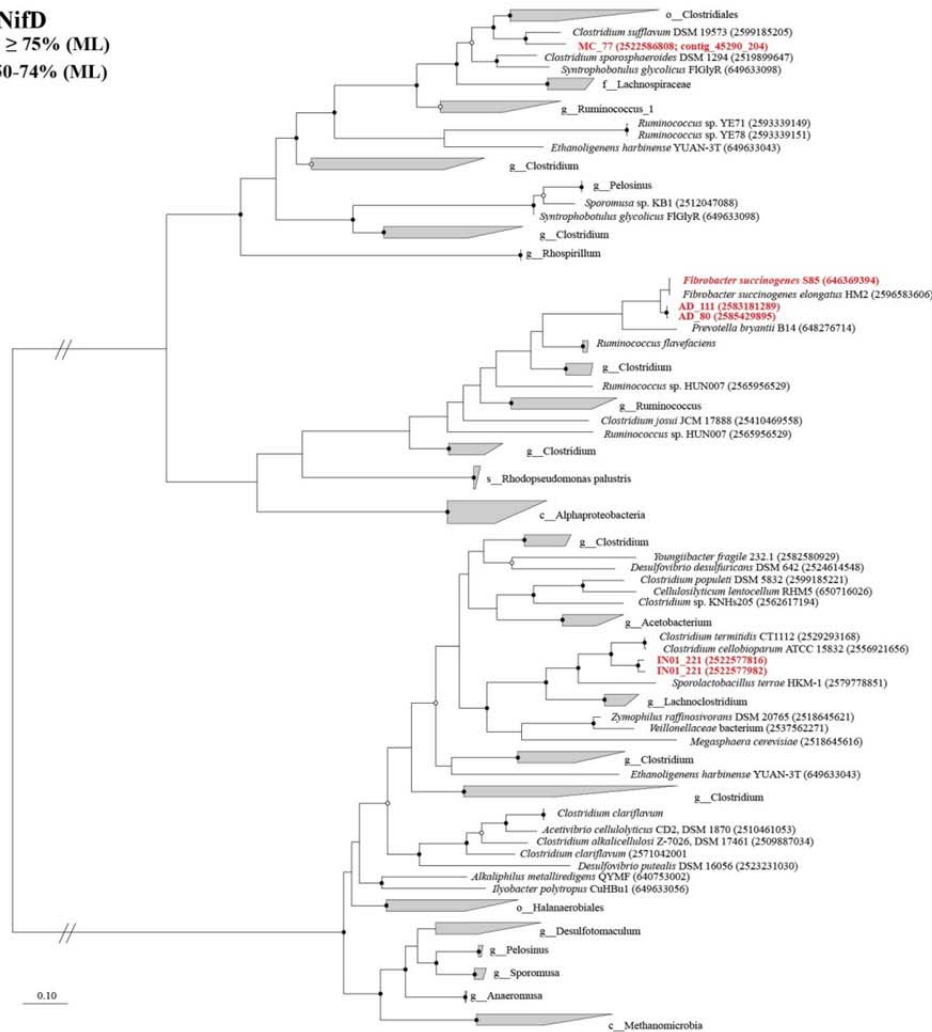


Figure S5.3: A maximum likelihood phylogenetic analysis of proteins encoding cytochrome bd-I ubiquinol oxidase; A) subunit 1 and B) subunit 2 > 341aa. Cytochrome bd-I ubiquinol oxidase protein sequences were aligned using MAFFT v7.221 and phylogenetic trees were constructed from 2,256 finished genomes from the IMG database (Markowitz et al., 2009) using Fasttree v2.1.7. The trees are unrooted and only the closest neighbors of the Fibrobacteres (in red) are shown with corresponding IMG IDs in brackets. Bootstrap support for interior nodes is indicated by dots according to the legend at the top left of the figure.

A. NifD

- $\geq 75\%$ (ML)
- 50-74% (ML)



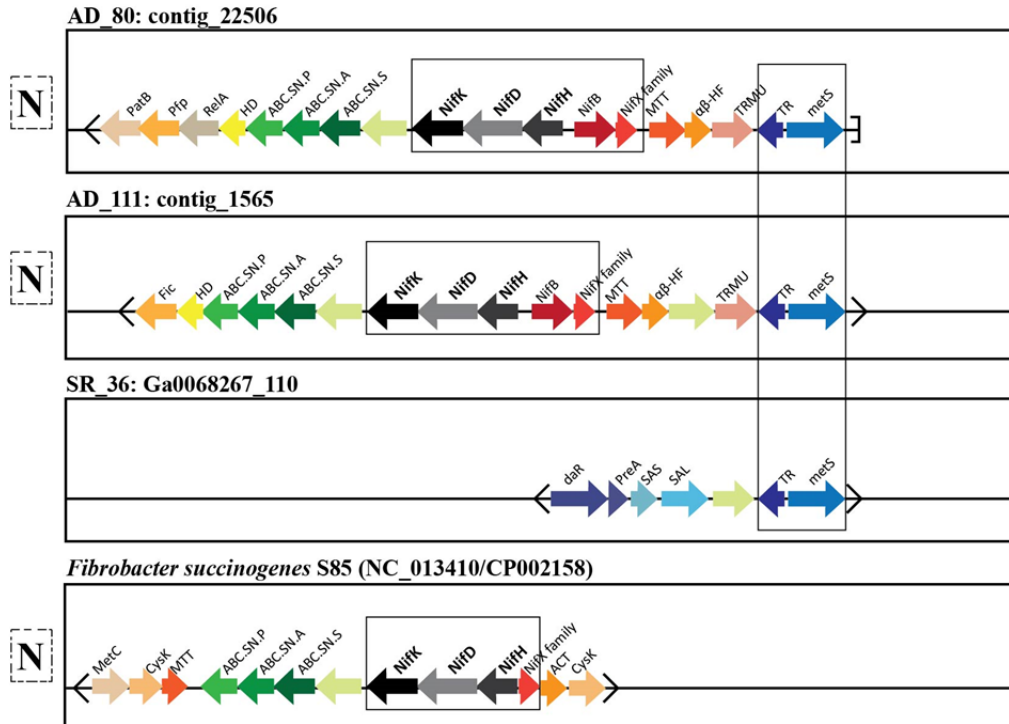
B. NifK

- $\geq 75\%$ (ML)
- 50-74% (ML)



Figure S5.4: A maximum likelihood phylogenetic analysis of protein encoding nitrogenase; A) NifD and B) NifK > 458aa. Nitrogenase protein sequences were aligned using MAFFT v7.221 and phylogenetic trees were constructed from 2,256 finished genomes from the IMG database (Markowitz et al., 2009) using Fasttree v2.1.7. The trees are unrooted and only the closest neighbors of the Fibrobacteres (in red) are shown with corresponding IMG IDs in brackets. Bootstrap support for interior nodes is indicated by dots according to the legend at the top left of the figure.

A. Fibrobacteraceae



B. Fibromonadaceae

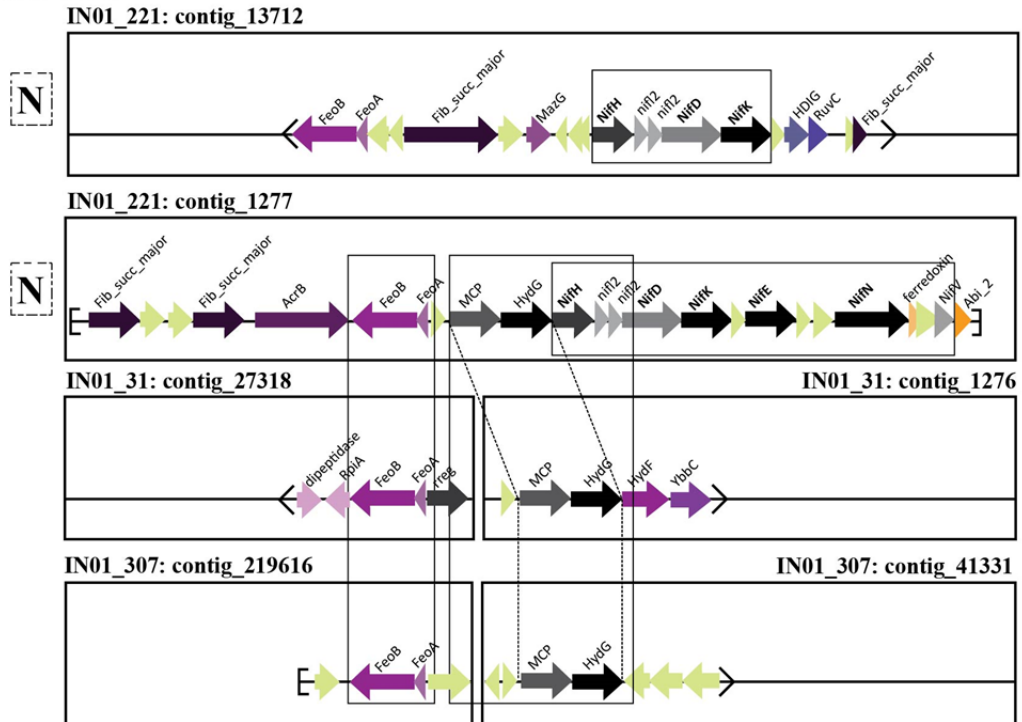


Figure S5.5: Gene neighborhoods of nitrogen fixing genes in the families A) Fibrobacteraceae and B) Fibromonadaceae. Genomes containing *nif* genes are indicated by dotted boxed Ns to the left of the figure. Colors indicate orthologous gene families and syntenous blocks of genes are highlighted by boxing; horizontal boxes denote *nif* genes and vertical boxes are common flanking region.

Families

- Fibrobacteraceae*
- Fibromonadaceae*
- C. alkaliphilus* AChtl
- MC_77

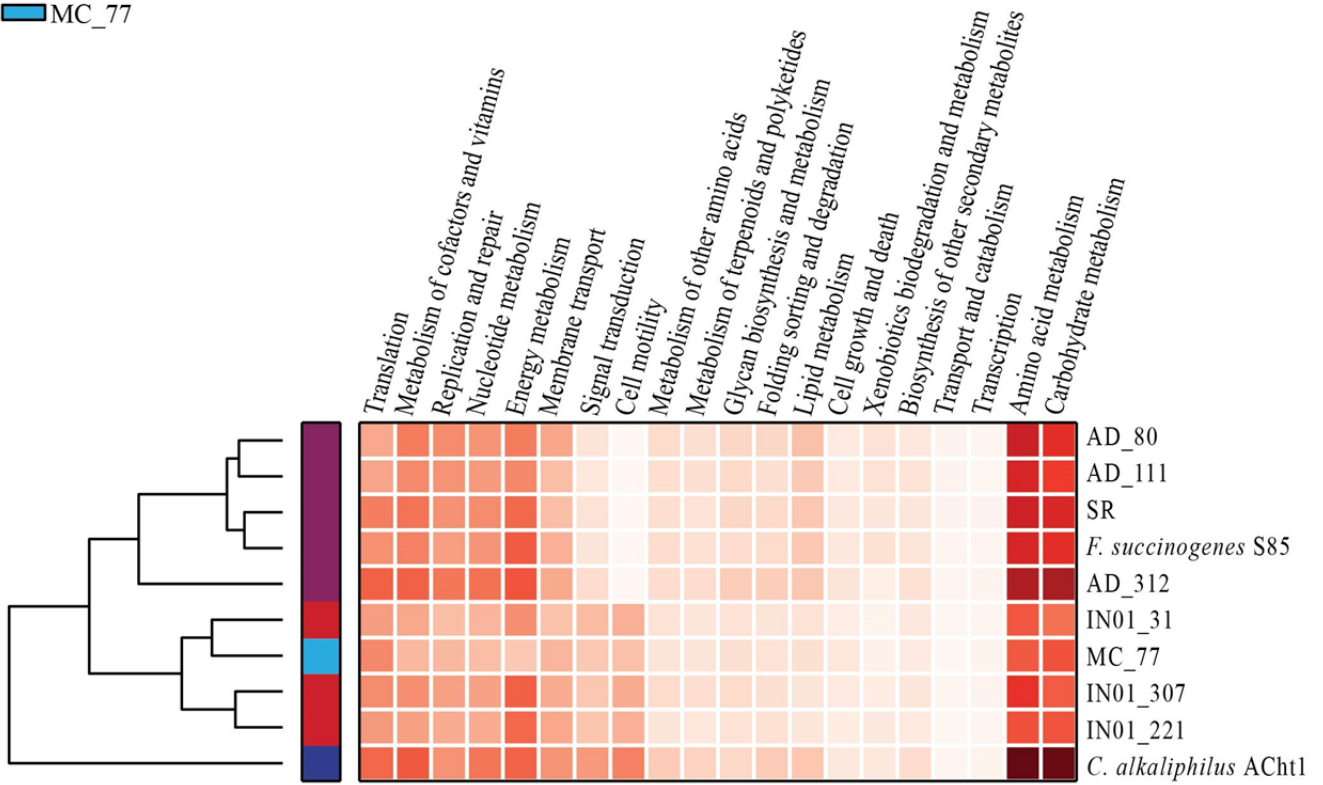
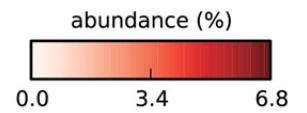
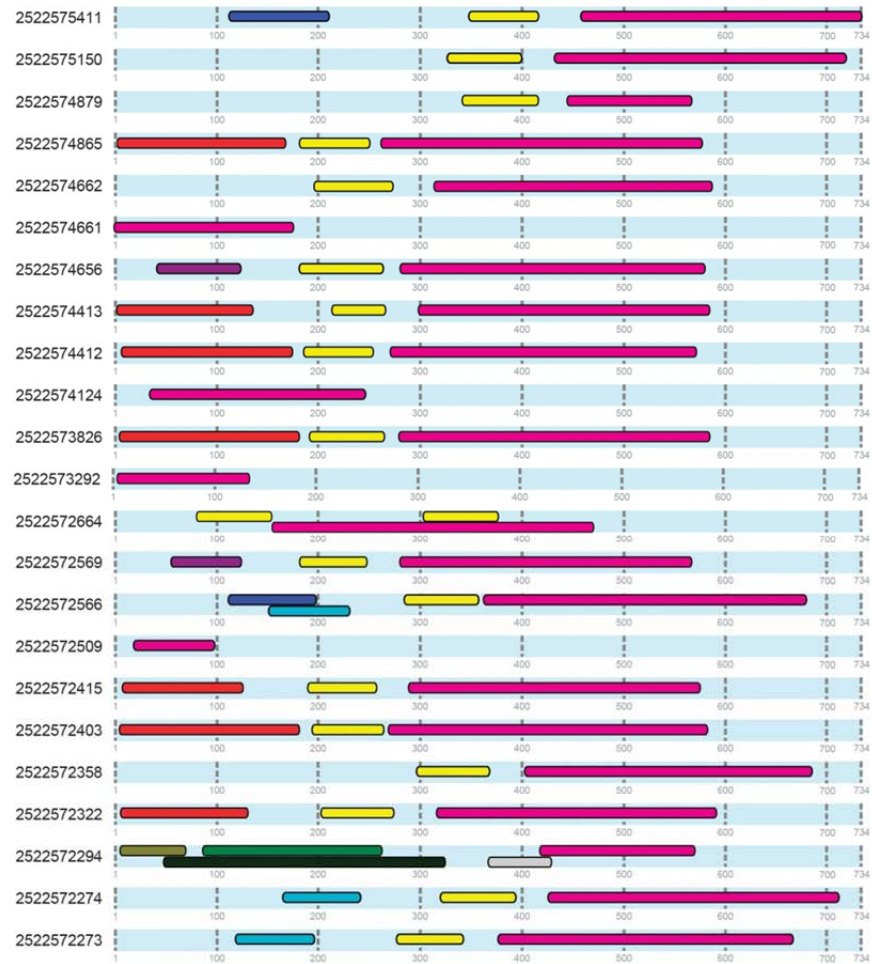


Figure S5.6: Relative abundance of KEGG ortholog (KO) functional categories across the investigated Fibrobacteres genomes. Genomes are ordered by similarity of relative abundance patterns.

A. IN01_31



Periplasmic sensor domain-like
 HAMP linker domain
 Methyl-accepting chemotaxis protein (MCP) signalling domain
 Fibrobacter succinogenes major paralogous domain
 4HB MCP domain
 Chemoreceptor zinc-binding domain

Cache domain
 4Fe-4S ferredoxin-type, iron-sulphur binding domain
 Iron hydrogenase, large subunit, C-terminal
 Iron hydrogenase
 4Fe-4S domain
 Target SNARE coiled-coil domain

B. IN01_221

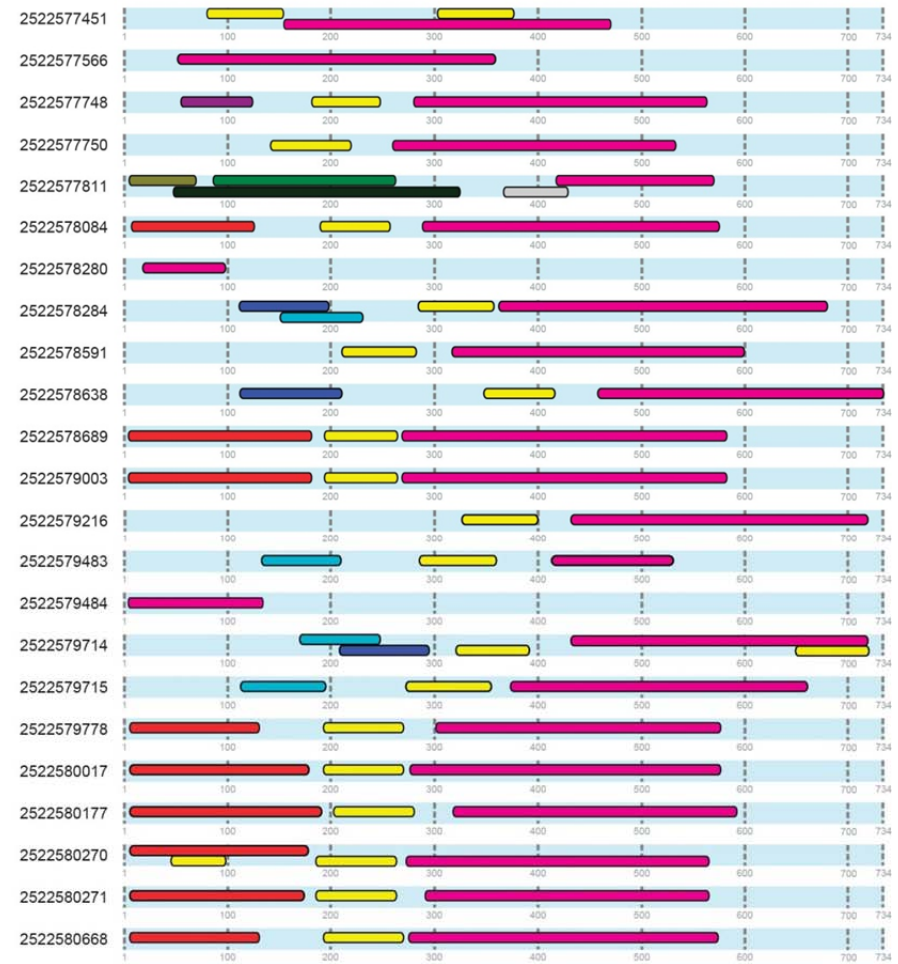
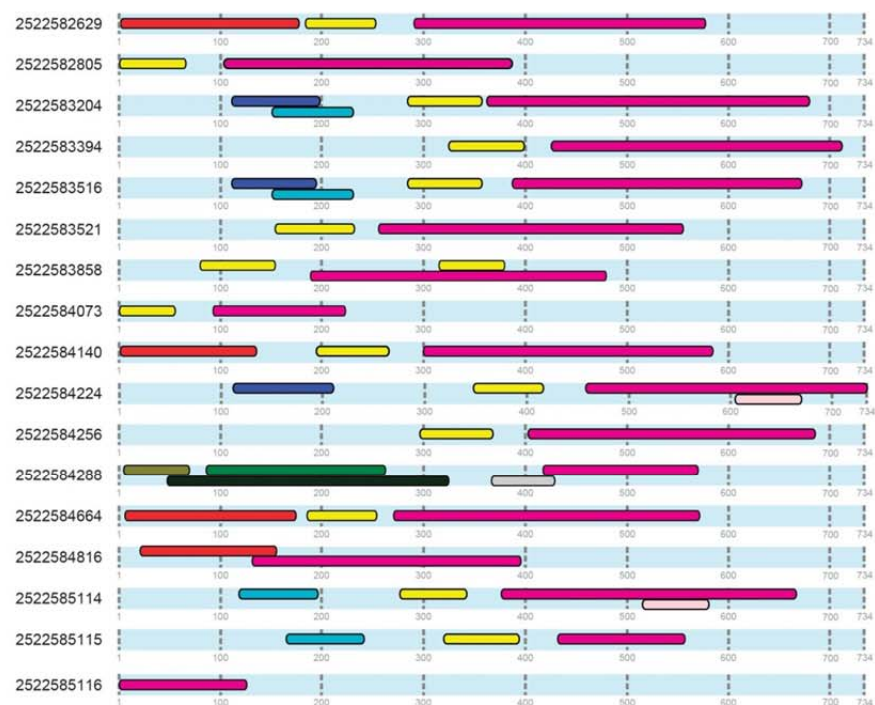


Figure S5.7: Methyl-accepting chemotaxis proteins (MCP) encoded in the genomes of family Fibromonadaceae (panels A-C) and class Chitinivibrionia (panels D-E) predicted with InterProScan5 (Jones et al., 2014). Protein domains are shown as colored blocks.

C. IN01_307



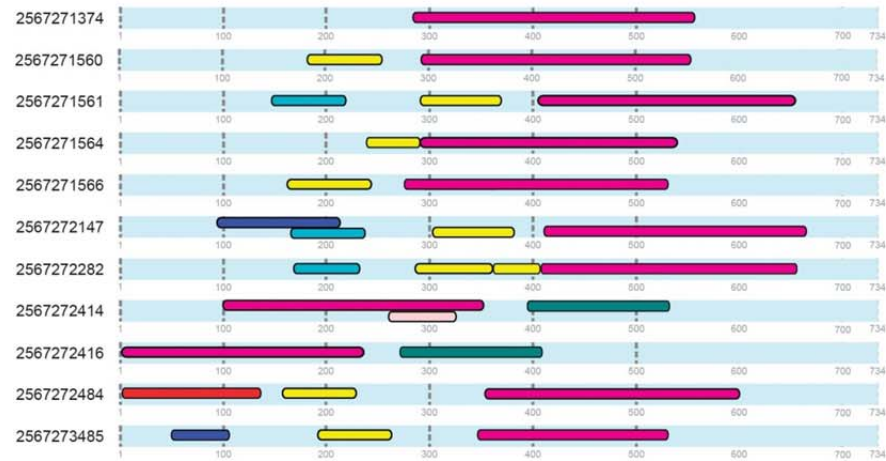
Periplasmic sensor domain-like
 HAMP linker domain
 Methyl-accepting chemotaxis protein (MCP) signalling domain
 Fibrobacter succinogenes major paralogous domain
 4HB MCP domain
 Chemoreceptor zinc-binding domain

Cache domain
 4Fe-4S ferredoxin-type, iron-sulphur binding domain
 Iron hydrogenase, large subunit, C-terminal
 Iron hydrogenase
 4Fe-4S domain
 Target SNARE coiled-coil domain

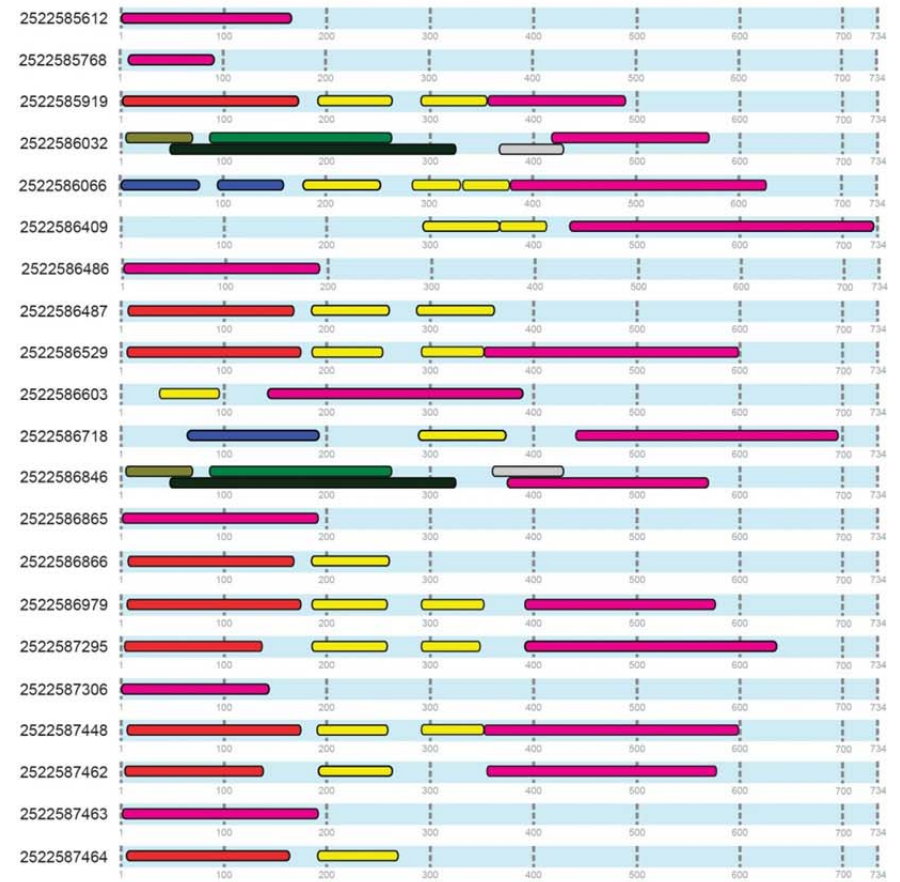
Cache domain
 Iron hydrogenase, large subunit, C-terminal
 Iron hydrogenase
 4Fe-4S domain
 Target SNARE coiled-coil domain

Figure S5.7: Continued.

D. *Chitinivibrio alkaliphilus* AChT1



E. MC_77



Periplasmic sensor domain-like
 HAMP linker domain
 Methyl-accepting chemotaxis protein (MCP) signalling domain
 Fibrobacter succinogenes major paralogous domain
 4HB MCP domain
 Chemoreceptor zinc-binding domain







 Cache domain
 4Fe-4S ferredoxin-type, iron-sulphur binding domain
 Iron hydrogenase, large subunit, C-terminal
 Iron hydrogenase
 4Fe-4S domain
 Target SNARE coiled-coil domain



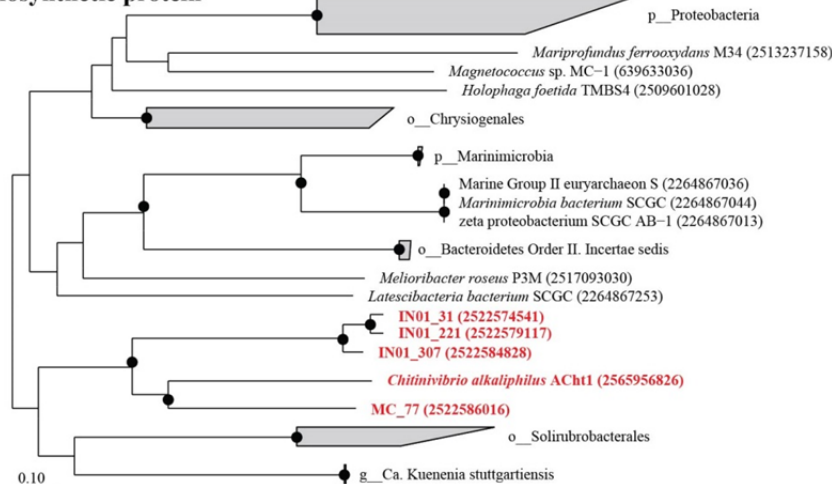





Figure S5.7: Continued.

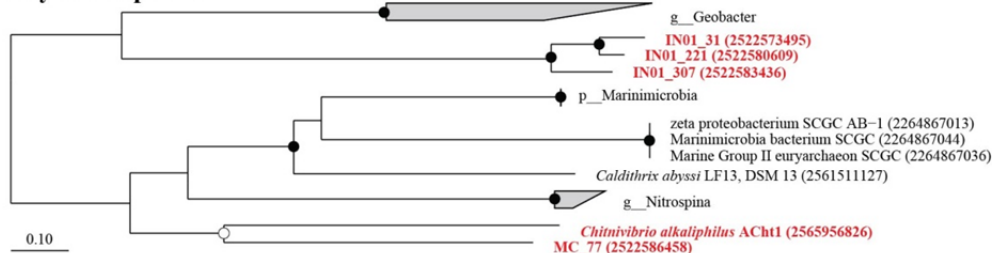
A. FlhA; flagellar biosynthetic protein

- ≥ 75% (ML)
- 50-74% (ML)



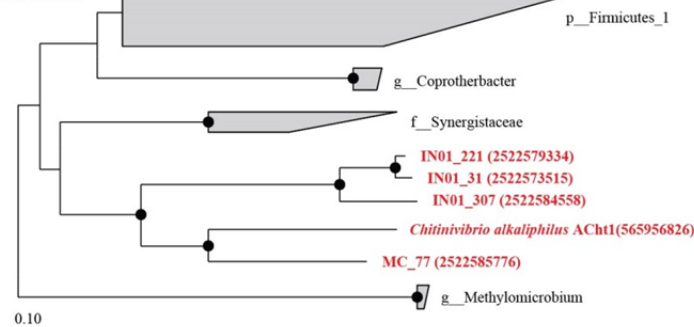
B. FlhB; flagellar biosynthetic protein

- ≥ 75% (ML)
- 50-74% (ML)



C. MotA; flagellar motor A

- ≥ 75% (ML)
- 50-74% (ML)



D. FliA; RNA polymerase sigma factor for flagellar

- ≥ 75% (ML)
- 50-74% (ML)

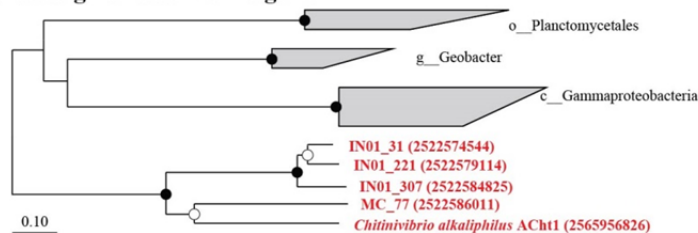
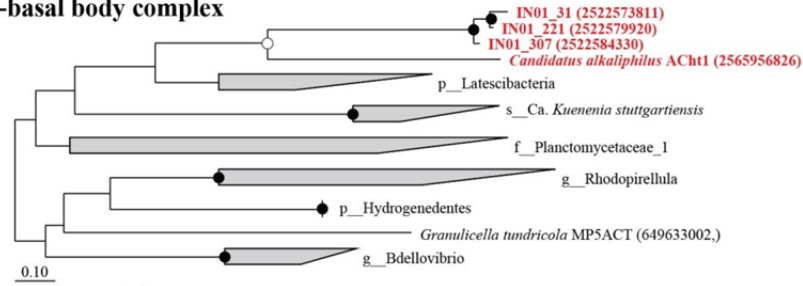


Figure S5.8: A maximum likelihood phylogenetic analysis of protein encoding flagellar genes; A) Flagellar biosynthetic proteins (FlhA), B) Flagellar motor proteins (FlhB), C) RNA polymerase sigma factor for flagellar (MotA), D) Flagellar hook-basal body complex (FliA). Flagellar protein sequences were aligned using MAFFT v7.221 and phylogenetic trees were constructed from 2,256 finished genomes from the IMG database (Markowitz et al., 2009) using Fasttree v2.1.7. The trees are unrooted and only the closest neighbors of the Fibrobacteres (in red) are shown with corresponding IMG IDs in brackets. Bootstrap support for interior nodes is indicated by dots according to the legend at the top left of the figure. The length of amino acid sequences ranges from 339 to 957.

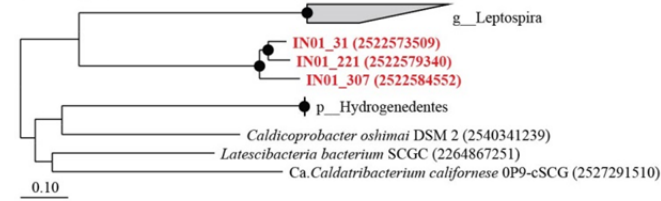
E. FliE; flagellar hook-basal body complex

- ≥ 75% (ML)
- 50-74% (ML)



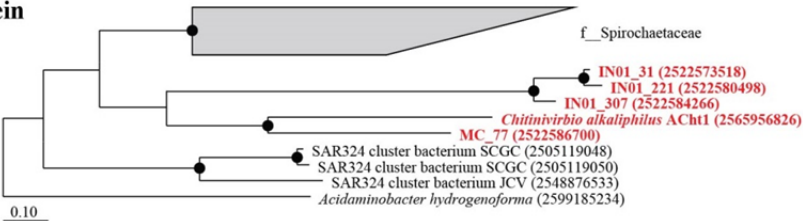
F. FliP; flagellar biosynthetic protein

- ≥ 75% (ML)
- 50-74% (ML)



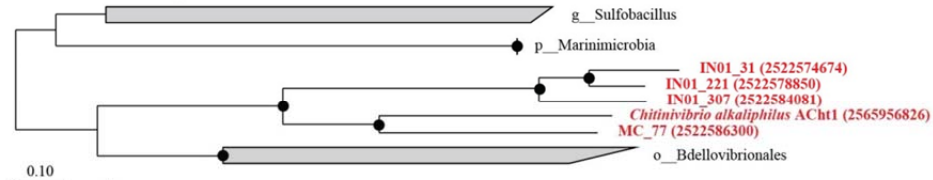
G. FliS; flagellar protein

- ≥ 75% (ML)
- 50-74% (ML)



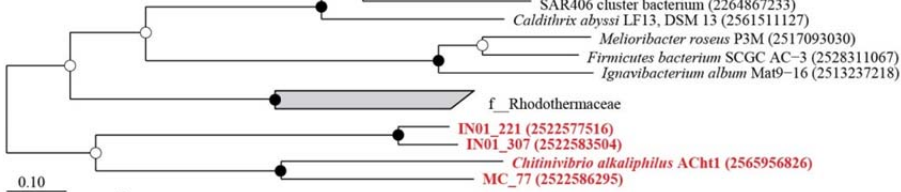
H. FlgK; flagellar hook-associated protein 1

- ≥ 75% (ML)
- 50-74% (ML)



I. FlgI; flagellar P-ring protein precursor

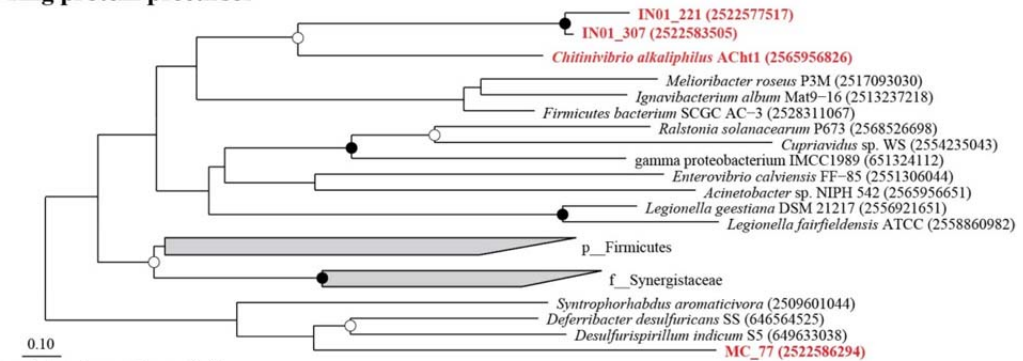
- ≥ 75% (ML)
- 50-74% (ML)



J. FlgH; flagellar L-ring protein precursor

Bootstrap values

- ≥ 75
- 50-74



K. FlgC; flagellar basal-body rod protein

Bootstrap values

- ≥ 75
- 50-74

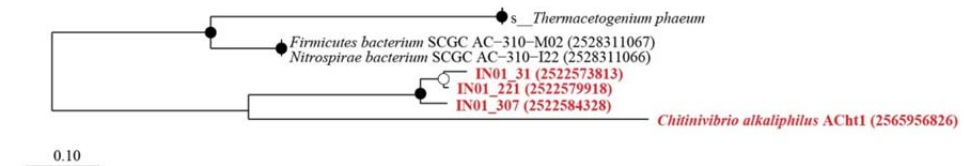


Figure S5.8: Continued.

Table S5.1: Metagenome sequencing statistics.

Sample ID	Shotgun sequence (Gb) (2x100bp)	Number of scaffolds ^a	Largest contig	N50a	Number of population genomes ^b	Habitat
IN01201	18.8	804,409	165,423	1,375	3 (65)	Termite gut
IN01202	17.7					Termite gut
IN01041	17.9					Termite gut
IN01042	19.1					Termite gut
MC05	25.3	607,574	89,930	1,084	1 (102)	Termite gut
MC06	26.6					Termite gut
MC07	19.2					Termite gut
AD1_T1	33.2	494,042	815,228	2,999	3 (100)	Anaerobic digester
AD2_T1	27.3					Anaerobic digester
AD3_T1	16.5					Anaerobic digester
AD1_T2	10.9					Anaerobic digester
AD2_T2	14.2					Anaerobic digester
AD3_T2	9.2					Anaerobic digester
SRR948090	9.9	78,410	423,553	2,425	1(36)	Sheep rumen

^a contigs \geq 500bp

^b Fibrobacteres (total)

Table S5.2: Shared orthologous genes and average amino acid identities (AAI) for all genome pairs, and proposed taxonomic rank relationships based on Kostantinidis & Tiedje, 2005.

Genome Id A	Genes in A	Genome Id B	Genes in B	No. orthologous genes	Mean AAI
<i>Intra Domain</i>					
<i>C. alkaliphilus</i> ACht1	2363	AD_80	2782	387	43.7
<i>C. alkaliphilus</i> ACht1	2363	AD_111	3039	385	44.1
<i>C. alkaliphilus</i> ACht1	2363	SR	2863	412	44.5
<i>C. alkaliphilus</i> ACht1	2363	<i>F. succinogenes</i> S85	3126	427	44.7
AD_80	2782	MC_77	2254	307	45.0
<i>C. alkaliphilus</i> ACht1	2363	IN01_307	2731	406	45.2
<i>C. alkaliphilus</i> ACht1	2363	IN01_221	3284	416	45.3
AD_111	3039	MC_77	2254	305	45.3
<i>C. alkaliphilus</i> ACht1	2363	IN01_31	3298	401	45.4
<i>C. alkaliphilus</i> ACht1	2363	AD_312	2362	427	45.6
<i>Intra Phylum</i>					
MC_77	2254	SR	2863	349	46.3
<i>F. succinogenes</i> S85	3126	MC_77	2254	345	46.5
AD_312	2362	MC_77	2254	319	47.2
IN01_307	2731	MC_77	2254	367	48.2
MC_77	2254	IN01_31	3298	373	48.7
IN01_221	3284	MC_77	2254	377	48.8
<i>C. alkaliphilus</i> ACht1	2363	MC_77	2254	499	49.0
<i>Intra Class</i>					
AD_111	3039	IN01_31	3298	781	53.7
AD_80	2782	IN01_31	3298	777	53.8
AD_80	2782	IN01_307	2731	761	53.9
AD_80	2782	IN01_221	3284	815	54.0
AD_111	3039	IN01_307	2731	762	54.1
AD_111	3039	IN01_221	3284	819	54.1
IN01_31	3298	SR	2863	842	54.3
<i>F. succinogenes</i> S85	3126	IN01_307	2731	845	54.4
<i>F. succinogenes</i> S85	3126	IN01_31	3298	845	54.4
<i>F. succinogenes</i> S85	3126	IN01_221	3284	886	54.5
IN01_307	2731	SR	2863	827	54.7
IN01_221	3284	SR	2863	883	54.8
AD_312	2362	IN01_221	3284	894	55.4
AD_312	2362	IN01_31	3298	849	55.5
AD_312	2362	IN01_307	2731	832	55.6
<i>Intra Order</i>					
AD_111	3039	AD_312	2362	1439	61.1
AD_312	2362	AD_80	2782	1392	61.3
AD_312	2362	SR	2863	1518	61.5
AD_312	2362	<i>F. succinogenes</i> S85	3126	1481	61.7
<i>Intra Family</i>					
AD_111	3039	<i>F. succinogenes</i> S85	3126	1903	73.1
AD_80	2782	<i>F. succinogenes</i> S85	3126	1877	73.2
<i>F. succinogenes</i> S85	3126	SR	2863	1921	73.5
<i>Intra Genus</i>					
AD_111	3039	SR	2863	1917	75.8
AD_80	2782	SR	2863	1851	76.3
IN01_307	2731	IN01_31	3298	1420	79.0
IN01_221	3284	IN01_307	2731	1520	79.9
IN01_221	3284	IN01_31	3298	1586	83.2
AD_111	3039	AD_80	2782	2003	84.9

Table S5.3: Presence/absence of genes in the Fibrobacteres genomes organised into functional pathways and modules depicted in Figure 2.

Functional description	Function ID	Symbol	Fibrobacteria							Chitinivibronia		
			Fibrobacteraceae				Fibromonadaceae			<i>C. alkaliphilus</i> Acht1	MC_77	
			<i>F. succinogenes</i> S85	AD_80	AD_111	SR_36	AD_312	IN01_31	IN01_221			IN01_307
Embden-Meyerhof-Parnas(glycolysis)												
Glucokinase	EC 2.7.1.2	GLK	1	1	1	1	1	1	1	-	1	1
Phosphoglucotomutase	EC 5.4.2.2	PGM	-	-	-	-	-	-	-	-	1	1
Glucose-6-phosphate isomerase	EC 5.3.1.9	G6P	1	1	1	1	1	-	1	-	1	1
6-phosphofruktokinase	EC 2.7.1.11	PFK	1	1	1	1	1	1	2	2	2	1
fructose-bisphosphate aldolase	EC 4.1.2.13	FBA	1	-	-	1	1	1	1	1	1	1
triosephosphate isomerase	EC 5.3.1.1	TPI	1	1	1	1	1	1	1	-	1	1
glyceraldehyde 3-phosphate dehydrogenase	EC 1.2.1.12	G3P	1	-	-	1	1	1	1	1	2	1
phosphoglycerate kinase	EC 2.7.2.3	PGK	1	1	1	1	1	1	1	1	1	-
phosphoglycerate mutase	EC 5.4.2.12	PGM	1	1	1	1	1	1	1	1	1	-
enolase	EC 4.2.1.11	-	1	1	1	-	1	1	1	1	1	-
pyruvate kinase	EC 2.7.1.40	PK	-	-	-	-	-	-	-	-	1	1
Chitin catabolic pathway												
chitinase [EC 3.2.1.14]	K01183	-	1	1	-	1	1	-	-	-	-	1
beta-N-acetylhexosaminidase [EC 3.2.1.52]	K01207	BNAH	2	2	2	1	1	-	-	-	-	-
beta-N-acetylglucosaminidase	pfam07555	BNAG	-	-	-	-	-	-	-	-	1	-
Cellobiose/Cellobiose phosphorylase	GH94	CBP	1	1	1	1	1	5	1	1	2	1
N-acetylglucosamine-1-P-mutase (Phosphomannomutase) [EC 5.4.2.8]	K01840	AGM	3	3	3	3	3	-	2	2	1	1
N-acetylglucosamine-6-phosphate deacetylase (Ribonucleotide monophosphatase NagD, HAD superfamily)	COG0647	NAG	2	2	2	2	2	1	1	1	2	-
Polysaccharide deacetylase	pfam01522	PD	3	1	1	1	1	2	2	2	9	-
exo-1,4-beta-D-glucosaminidase	K15855	BGA	-	-	-	-	-	-	-	-	1	-
glucokinase [EC 2.7.1.2]	K00845	GLK	1	1	1	1	1	1	1	-	1	1
glucosamine-6-phosphate deaminase [EC 3.5.99.6]	K02564	G6PD	-	-	-	-	-	-	-	-	1	-
Pyruvate metabolism												
Phosphoenolpyruvate carboxykinase [EC 4.1.1.32]	pfam00821	PEPC	1	1	2	1	1	1	2	1	-	-
phosphoenolpyruvate carboxylase [EC 4.1.1.31]	K01595	PEPC	-	-	-	-	-	-	-	-	1	-
oxaloacetate decarboxylase, alpha subunit [EC 4.1.1.3]	K01571	-	-	-	-	-	-	-	-	-	1	1
oxaloacetate decarboxylase, beta subunit [EC 4.1.1.3]	K01572	-	1	1	1	1	1	1	1	1	1	-
pyruvate carboxylase subunit A [EC 6.4.1.1]	K01959	-	-	-	-	-	-	1	1	-	-	-
pyruvate carboxylase subunit B [EC 6.4.1.1]	K01960	-	1	1	1	1	1	1	-	1	1	-
malate dehydrogenase [EC 1.1.1.37]	K00024	-	1	-	-	1	1	1	1	1	-	-
malate dehydrogenase (oxaloacetate-decarboxylating)(NADP+) [EC 1.1.1.40]	K00029	-	1	1	1	1	1	-	-	-	-	-
fumarate hydratase, class I [EC 4.2.1.2]	K01676	-	1	1	1	1	1	1	1	1	-	-
fumarate hydratase subunit alpha [EC 4.2.1.2]	K01677	-	1	-	-	-	-	-	-	-	1	1
fumarate hydratase subunit beta [EC 4.2.1.2]	K01678	-	1	-	-	-	-	-	-	-	1	1
formate C-acetyltransferase [EC 2.3.1.54]	K00656	FCA	1	1	1	1	1	-	-	-	1	-
pyruvate formate lyase activating enzyme [EC 1.97.1.4]	K04069	-	1	1	1	1	2	1	1	-	3	2
putative pyruvate-flavodoxin oxidoreductase [EC 1.2.7.-]	K03737	-	1	-	-	1	1	1	1	1	1	-
acetyl-CoA synthetase [EC 6.2.1.1]	K01895	ACS	1	1	1	1	1	1	-	-	-	-
phosphate acetyltransferase [EC 2.3.1.8]	K00625	PAT	-	-	-	-	-	-	-	-	1	-
acetate kinase [EC 2.7.2.1]	K00925	AK	1	1	1	1	1	-	1	1	1	-
acetaldehyde dehydrogenase / alcohol dehydrogenase [EC 1.2.1.10 1.1.1.1]	K04072	AAD	-	-	-	-	-	-	-	-	1	-
aldehyde dehydrogenase (NAD+) [EC 1.2.1.3]	K00128	ADH	1	1	1	1	1	-	-	-	1	-
acylphosphatase [EC 3.6.1.7]	K01512	-	-	-	-	-	-	-	-	-	1	1
pyruvate, orthophosphate dikinase [EC 2.7.9.1]	K01006	-	-	-	-	-	1	1	1	1	1	-
pyruvate, water dikinase [EC 2.7.9.2]	K01007	-	-	-	-	-	-	-	-	-	1	-

Table S5.3: Continued.

Functional description	Function ID	Symbol	Fibrobacteria							Chitinivibronia		
			Fibrobacteraceae				Fibromonadaceae			<i>C. alkaliphilus</i> AChT1	MC_77	
			<i>F. succinogenes</i> S85	AD_80	AD_111	SR_36	AD_312	IN01_31	IN01_221			IN01_307
Pentose phosphate pathway												
glucose-6-phosphate 1-dehydrogenase	EC 1.1.1.49	-	-	-	-	-	-	-	-	-	1	-
6-phosphogluconolactonase	EC 3.1.1.31	-	-	-	-	-	-	-	-	-	1	-
6-phosphogluconate dehydrogenase	EC 1.1.1.44	-	-	-	-	-	-	-	-	-	1	-
ribulose-phosphate 3-epimerase	EC 5.1.3.1	RPE	1	1	1	1	1	1	1	2	1	1
ribose 5-phosphate isomerase	EC 5.3.1.6	RPI	1	-	1	1	1	1	1	1	1	-
transketolase	EC 2.2.1.1	-	1	1	1	1	1	1	1	1	1	1
transaldolase	EC 2.2.1.2	-	-	-	-	-	-	-	-	-	1	-
ribose-phosphate pyrophosphokinase	EC 2.7.6.1	RPP	1	1	1	1	1	1	1	1	1	1
Tricarboxylic acid cycle (TCA)/Reverse TCA cycle												
Citrate (S)-synthase	EC 2.3.3.1	CS	1	1	1	1	2	1	2	1	1	1
Aconitate hydratase	EC 4.2.1.3	ACH	1	1	1	1	1	1	1	1	2	1
Isocitrate dehydrogenase (NADP(+))	EC 1.1.1.42	IDH	1	1	1	1	1	1	1	1	1	1
Isocitrate dehydrogenase (NAD(+))	EC 1.1.1.41	IDH	-	-	-	-	-	-	-	-	1	-
2-oxoglutarate dehydrogenase	EC 1.2.4.2	-	-	-	-	-	-	-	-	-	-	-
Dihydrolipoyllysine-residue succinyltransferase	EC 2.3.1.61	-	-	-	-	-	-	-	-	-	-	-
2-oxoglutarate synthase	EC 1.2.7.3	2OS	-	-	-	-	-	-	1	-	2	2
Succinyl-CoA-synthase	EC 6.2.1.5	SCS	-	-	-	-	-	-	-	-	-	-
Succinate dehydrogenase (quinone)	EC 1.3.5.1	SDH	3	1	2	2	2	2	2	2	2	2
Fumarate/Fumarate hydratase	EC 4.2.1.2	FH	3	1	1	1	1	1	1	1	2	2
Malate dehydrogenase	EC 1.1.1.37	MDH	1	-	-	1	1	1	1	1	-	-
Malic enzyme	COG0281	ME	1	1	1	1	1	-	-	-	1	1
Fumarate reductase (quinol)	EC 1.3.1.6	-	-	-	-	-	-	-	-	-	-	-
Glycoxylate cycle												
Malate synthase	EC 2.3.3.9	-	-	-	-	-	-	-	-	-	-	-
Isocitrate lyase	EC 4.1.3.1	-	-	-	-	-	-	-	-	-	-	-
<i>fermentative product</i>			<i>succinate</i>							<i>fumarate</i>		
Electron Transport Chain												
succinate dehydrogenase flavoprotein subunit [EC 1.3.5.1]	K00239	-	1	-	1	1	1	1	1	1	-	-
succinate dehydrogenase iron-sulfur subunit [EC 1.3.5.1]	K00240	-	2	1	1	1	1	1	1	1	-	-
succinate dehydrogenase cytochrome b556 subunit	K00241	-	1	-	1	1	1	1	1	1	-	-
NADH-quinone oxidoreductase subunit A [EC 1.6.5.3]	K00330	-	2	2	2	2	2	2	2	2	-	-
NADH-quinone oxidoreductase subunit B [EC 1.6.5.3]	K00331	-	1	1	1	1	1	1	1	1	-	-
NADH-quinone oxidoreductase subunit C [EC 1.6.5.3]	K00332	-	1	1	1	1	1	1	1	1	-	-
NADH-quinone oxidoreductase subunit D [EC 1.6.5.3]	K00333	-	2	2	2	2	2	2	3	2	-	-
NADH-quinone oxidoreductase subunit E [EC 1.6.5.3]	K00334	-	-	-	-	-	-	-	-	-	2	-
NADH-quinone oxidoreductase subunit F [EC 1.6.5.3]	K00335	-	1	1	1	1	1	1	-	-	1	-
NADH-quinone oxidoreductase subunit G [EC 1.6.5.3]	K00336	-	1	1	1	1	1	2	-	-	1	-
NADH-quinone oxidoreductase subunit H [EC 1.6.5.3]	K00337	-	2	2	2	2	2	1	2	2	-	-
NADH-quinone oxidoreductase subunit I [EC 1.6.5.3]	K00338	-	2	2	2	2	2	1	2	2	-	-
NADH-quinone oxidoreductase subunit J [EC 1.6.5.3]	K00339	-	2	2	2	2	2	1	2	2	-	-
NADH-quinone oxidoreductase subunit K [EC 1.6.5.3]	K00340	-	2	2	2	2	2	1	2	2	-	-
NADH-quinone oxidoreductase subunit L [EC 1.6.5.3]	K00341	-	2	2	2	2	2	2	2	2	-	-
NADH-quinone oxidoreductase subunit M [EC 1.6.5.3]	K00342	-	2	2	2	2	2	2	2	2	-	-
NADH-quinone oxidoreductase subunit N [EC 1.6.5.3]	K00343	-	2	2	2	2	2	2	2	2	-	-
NADH dehydrogenase [EC 1.6.99.3]	K00356	-	1	1	1	1	1	-	-	-	-	-
cytochrome d ubiquinol oxidase subunit I [EC 1.10.3.-]	K00425	-	-	-	-	-	-	1	1	-	-	-
cytochrome d ubiquinol oxidase subunit II [EC 1.10.3.-]	K00426	-	-	-	-	-	-	1	1	-	-	-

Table S5.3: Continued.

Functional description	Function ID	Symbol	Fibrobacteria							Chitinivibronia			
			Fibrobacteraceae				Fibromonadaceae			<i>C. alkaliphilus</i> ACh1	MC_77		
			<i>F. succinogenes</i> S85	AD_80	AD_111	SR_36	AD_312	IN01_31	IN01_221			IN01_307	
Electron Transport Chain													
ABC-type transport system involved in cytochrome bd biosynthesis, fused ATPase and permease components	COG4987	-	-	-	-	-	-	-	2	2	2	-	-
ABC-type transport system involved in cytochrome bd biosynthesis, ATPase and permease components	COG4988	-	-	-	-	-	-	-	1	1	1	-	-
polyphosphate kinase [EC 2.7.4.1]	K00937	-	-	1	-	-	-	-	-	1	1	1	1
inorganic pyrophosphatase [EC 3.6.1.1]	K01507	-	1	1	-	1	1	1	-	-	-	1	-
V/A-type H ⁺ -transporting ATPase subunit A [EC 3.6.3.14]	K02117	-	1	1	1	1	1	1	-	1	1	1	-
V/A-type H ⁺ -transporting ATPase subunit B	K02118	-	1	-	-	1	1	1	1	1	1	1	-
V/A-type H ⁺ -transporting ATPase subunit D	K02120	-	1	-	-	1	1	1	1	1	1	1	-
V/A-type H ⁺ -transporting ATPase subunit E	K02121	-	1	1	1	1	1	1	-	1	1	1	-
V/A-type H ⁺ -transporting ATPase subunit I	K02123	-	1	-	1	1	1	1	1	1	1	1	-
V/A-type H ⁺ -transporting ATPase subunit K	K02124	-	2	1	1	2	1	1	1	1	1	1	-
NADH dehydrogenase [EC 1.6.99.3]	K03885	-	-	-	-	-	-	-	-	-	-	1	-
NADH dehydrogenase (ubiquinone) flavoprotein 2 [EC 1.6.5.3 1.6.99.3]	K03943	-	-	-	-	-	-	-	1	-	1	-	-
NAD(P)H-quinone oxidoreductase subunit K [EC 1.6.5.3]	K05582	-	-	-	-	-	-	-	1	-	-	-	-
pyrophosphatase PpaX [EC 3.6.1.1]	K06019	-	-	-	-	-	-	-	-	-	-	1	-
manganese-dependent inorganic pyrophosphatase [EC 3.6.1.1]	K15986	-	1	1	-	-	-	-	-	-	-	-	-
K(+)-stimulated pyrophosphate-energized sodium pump [EC 3.6.1.1]	K15987	-	-	-	1	1	1	1	1	1	1	-	-
Galactose metabolism													
alpha-glucosidase [EC 3.2.1.20]	K01187	-	-	-	-	-	-	-	-	1	1	-	-
beta-galactosidase [EC 3.2.1.23]	K01190	-	1	1	1	1	1	-	-	-	-	-	-
UDP-glucose 4-epimerase [EC 5.1.3.2]	K01784	-	3	2	3	2	2	4	2	2	2	1	1
UDP-galactopyranose mutase [EC 5.4.99.9]	K01854	-	1	-	-	-	-	1	-	-	1	-	-
alpha-galactosidase [EC 3.2.1.22]	K07407	-	1	1	1	1	1	-	-	-	-	-	-
UDP-glucose 4-epimerase [EC 5.1.3.2]	K17716	-	2	-	-	-	-	-	1	-	-	-	-
Xylose metabolism													
D-xylose reductase	K17743	XR	-	-	-	-	-	-	-	-	-	-	-
Aldehyde reductase	K00011	AR	-	-	-	-	-	-	-	-	-	-	-
D-xylose reductase	K03331	XLR	-	-	-	-	-	-	-	-	-	-	-
L-iditol 2-dehydrogenase	K00008	IDH	-	-	-	-	-	-	-	-	-	-	-
xylose isomerase	K01805	XI	-	-	-	-	-	-	-	-	-	-	-
xylokinae	K00854	XLK	1	-	-	-	-	-	-	-	-	-	-
Nitrogen fixation													
nitrogen fixation protein NifB	K02585	-	-	-	-	-	-	-	-	-	-	-	-
nitrogenase molybdenum-iron protein alpha chain nifD	K02586	-	2	1	1	-	-	-	-	2	-	-	1
nitrogenase molybdenum-cofactor synthesis protein nifE	K02587	-	-	-	-	-	-	-	-	1	-	-	1
nitrogenase iron protein nifH	K02588	-	2	1	1	-	-	-	-	2	-	-	1
nitrogen regulatory protein PII 1 nifDH1	K02589	-	-	-	-	-	-	-	-	2	-	-	-
nitrogen regulatory protein PII 1 nifDH1	K02590	-	-	-	-	-	-	-	-	2	-	-	-
nitrogenase molybdenum-iron protein beta chain nifK	K02591	-	2	1	1	-	-	-	-	2	-	-	1
nitrogenase molybdenum-iron protein nifN	K02592	-	-	-	-	-	-	-	-	1	-	-	-
Ammonia uptake and assimilation													
Ammonia channel protein AmtB (permease)	COG0004	-	1	1	1	1	1	-	1	1	1	-	1
Nitrogen regulatory protein PII	COG0347	-	1	1	1	1	1	-	6	1	1	-	1
Glutamate synthase domain 2	COG0069	GS	1	1	1	1	1	1	1	1	1	-	-
Glutamine synthetase	COG0174	GS	-	-	-	1	-	-	1	-	1	-	-

Table S5.3: Continued.

Functional description	Function ID	Symbol	Fibrobacteria							Chitinivibronia		
			Fibrobacteraceae				Fibromonadaceae			<i>C. alkaliphilus</i> ACh1	MC_77	
			<i>F. succinogenes</i> S85	AD_80	AD_111	SR_36	AD_312	IN01_31	IN01_221			IN01_307
Ammonia uptake and assimilation												
Glutamine synthetase type III	COG3968	GS	2	2	2	2	2	1	1	1	2	1
Glutamate dehydrogenase/leucine dehydrogenase	COG0334	GD	1	1	1	1	1	-	1	1	2	-
Histidine ammonia-lyase	COG2986	-	1	1	-	-	-	1	-	-	-	-
Carbamoylphosphate synthase large subunit	COG0458	-	3	1	2	1	2	1	1	1	1	-
Sulfur metabolism												
<i>Sulfate transport system</i>												
sulfate transport system ATP-binding protein	K02045	-	1	-	1	1	-	1	-	-	-	-
sulfate transport system permease protein	K02046	-	1	-	1	1	-	1	-	1	-	-
sulfate transport system permease protein	K02047	-	1	-	1	1	-	1	1	1	-	-
sulfate transport system substrate-binding protein	K02048	-	2	-	1	1	-	1	1	1	-	-
<i>Assimilatory sulfate reduction (sulfate->sulfite)</i>												
bifunctional enzyme CysN/CysC	K00955	-	-	-	-	-	-	-	-	-	1	-
sulfate adenylyltransferase subunit 1	K00956	-	1	-	1	1	-	1	1	1	-	-
sulfate adenylyltransferase subunit 2	K00957	-	1	-	1	1	-	1	1	1	1	-
3'(2'), 5'-bisphosphate nucleotidase	K01082	-	-	-	-	-	1	-	-	-	-	-
phosphoadenosine phosphosulfate reductase	K00390	-	1	-	-	-	-	-	-	-	1	-
adenylylsulfate reductase, subunit A	K00394	-	1	-	1	1	-	1	-	1	-	-
adenylylsulfate reductase, subunit B	K00395	-	1	-	1	1	-	1	-	1	-	-
<i>ABC transport system</i>												
sulfonate transport system ATP-binding protein	K15555	-	1	1	1	-	-	-	-	-	-	-
Hydrogenase												
Iron only hydrogenase large subunit, C-terminal domain	COG4624	-	-	-	-	-	-	1	-	-	3	-
Iron only hydrogenase large subunit, C-terminal domain	pfam02906	-	-	-	-	-	-	3	1	1	3	2
Flagellar assembly (core genes are bolded)												
flagella basal body P-ring formation protein FlgA	K02386	-	-	-	-	-	-	1	1	1	1	-
flagellar basal-body rod protein FlgB	K02387	-	-	-	-	-	-	2	2	2	1	1
flagellar basal-body rod protein FlgC	K02388	-	-	-	-	-	-	2	2	2	1	-
flagellar basal-body rod modification protein FlgD	K02389	-	-	-	-	-	-	1	1	1	1	-
flagellar hook protein FlgE	K02390	-	-	-	-	-	-	2	2	2	1	1
flagellar basal-body rod protein FlgF	K02391	-	-	-	-	-	-	-	-	-	-	-
flagellar basal-body rod protein FlgG	K02392	-	-	-	-	-	-	3	2	1	2	-
flagellar L-ring protein precursor FlgH	K02393	-	-	-	-	-	-	-	1	1	1	1
flagellar P-ring protein precursor FlgI	K02394	-	-	-	-	-	-	-	1	1	1	1
flagellar hook-associated protein 1 FlgK	K02396	-	-	-	-	-	-	1	1	1	1	1
flagellar hook-associated protein 3 FlgL	K02397	-	-	-	-	-	-	1	1	1	1	-
negative regulator of flagellin synthesis FlgM	K02398	-	-	-	-	-	-	-	-	-	-	-
flagellar biosynthesis protein FlhA	K02400	-	-	-	-	-	-	1	1	1	1	1
flagellar biosynthetic protein FlhB	K02401	-	-	-	-	-	-	1	1	1	1	1
flagellar transcriptional activator FlhC	K02402	-	-	-	-	-	-	-	-	-	-	-
flagellar transcriptional activator FlhD	K02403	-	-	-	-	-	-	-	-	-	-	-
flagellin, FlhC	K02406	-	-	-	-	-	-	7	5	3	5	3
flagellar hook-associated protein 2, FlhD	K02407	-	-	-	-	-	-	1	1	2	2	1
flagellar hook-basal body complex protein FlhE	K02408	-	-	-	-	-	-	1	1	1	1	-
flagellar M-ring protein FlhF	K02409	-	-	-	-	-	-	1	1	1	1	-
flagellar motor switch protein FlhG	K02410	-	-	-	-	-	-	1	3	2	1	-
flagellar assembly protein FlhH	K02411	-	-	-	-	-	-	1	1	1	1	-
flagellum-specific ATP synthase [EC:3.6.3.14]	K02412	-	-	-	-	-	-	1	1	1	1	-
flagellar FlhJ protein	K02413	-	-	-	-	-	-	1	1	1	1	-
flagellar hook-length control protein FlhK	K02414	-	-	-	-	-	-	1	-	-	1	-

Table S5.3: Continued.

Functional description	Function ID	Symbol	Fibrobacteria							Chitinivibronia			
			<i>F. succinogenes</i> S85	Fibrobacteraceae			Fibromonadaceae			<i>C. alkaliphilus</i> AChT1	MC_77		
				AD_80	AD_111	SR_36	AD_312	IN01_31	IN01_221			IN01_307	
Flagellar assembly													
flagellar motor switch protein FlIM	K02416	-	-	-	-	-	-	-	1	1	1	1	1
flagellar motor switch protein FlIN/FliY	K02417	-	-	-	-	-	-	-	2	2	2	2	2
flagellar protein FlIO/FlIZ	K02418	-	-	-	-	-	-	-	-	-	-	-	-
flagellar biosynthetic protein FlIP	K02419	-	-	-	-	-	-	-	1	1	1	-	-
flagellar biosynthetic protein FlIQ	K02420	-	-	-	-	-	-	-	1	1	1	1	-
flagellar biosynthetic protein FlIR	K02421	-	-	-	-	-	-	-	1	1	1	1	-
flagellar protein FlIS	K02422	-	-	-	-	-	-	-	1	1	1	1	1
flagellar protein FlIT	K02423	-	-	-	-	-	-	-	-	-	-	-	-
flagellar motor A MotA	K02556	-	-	-	-	-	-	-	1	1	1	1	1
flagellar motor B MotB	K02557	-	-	-	-	-	-	-	1	1	1	1	1
flagella synthesis protein FlgN	K20399	-	-	-	-	-	-	-	-	-	-	-	-
Bacterial chemotaxis													
methyl-accepting chemotaxis protein	K03406	MCP	-	-	-	-	-	-	14	18	12	8	11
chemotaxis protein CheD	K03411	cheD	-	-	-	-	-	-	-	-	-	3	2
chemotaxis protein methyltransferase CheR	K00575	cheR	-	-	-	-	-	-	1	3	1	2	1
two-component system, chemotaxis family, response regulator CheB	K03412	cheB	-	-	-	-	-	-	1	1	1	2	1
two-component system, chemotaxis family, sensor kinase CheA	K03407	cheA	-	-	-	-	-	-	2	1	1	5	1
purine-binding chemotaxis protein CheW	K03408	cheW	-	-	-	-	-	-	1	1	1	3	2
two-component system, chemotaxis family, response regulator CheV	K03415	cheV	-	-	-	-	-	-	-	-	-	1	-
two-component system, chemotaxis family, response regulator CheY	K03413	cheY	-	-	-	-	1	-	2	2	1	9	2
chemotaxis protein CheC	K03410	cheC	-	-	-	-	-	-	-	-	-	-	-
chemotaxis protein CheZ	K03414	cheZ	-	-	-	-	-	-	-	-	-	-	-
chemotaxis protein CheX	K03409	cheX	-	-	-	-	-	-	2	2	3	1	2
Amino acid													
Glycine, serine and threonine metabolism													
<i>Serine biosynthesis (glycerate 3P->serine)</i>													
D-3-phosphoglycerate dehydrogenase [EC:1.1.1.95]	K00058	-	2	1	1	1	1	1	1	1	1	1	1
phosphoserine aminotransferase [EC:2.6.1.52]	K00831	-	1	1	1	1	1	1	1	1	1	1	1
phosphoserine / homoserine phosphotransferase [EC:3.1.3.3 2.7.1.39]	EC 3.1.3.3 EC 2.7.3.9	-	1	1	1	1	1	2	1	1	1	-	-
<i>Glycine biosynthesis (serine->glycine)</i>													
glycine hydroxymethyltransferase [EC:2.1.2.1]	K00600	-	1	1	1	1	1	1	1	1	1	1	-
<i>Threonine biosynthesis (aspartate->homoserine->threonine)</i>													
aspartate kinase [EC:2.7.2.4]	K00928	-	1	1	1	1	1	1	1	1	1	1	1
aspartate-semialdehyde dehydrogenase [EC:1.2.1.11]	K00133	-	1	1	1	1	1	1	1	1	1	1	2
homoserine dehydrogenase [EC:1.1.1.3]	K00003	-	1	1	1	1	1	1	1	1	1	1	1
homoserine kinase type II [EC:2.7.1.39]	K02204	-	1	1	1	1	1	2	1	1	1	1	-
threonine synthase [EC:4.2.3.1]	K01733	-	1	1	1	1	1	1	1	1	1	1	1
Valine, Isoleucine, leucine biosynthesis (Pyruvate -> Valine/Isoleucine/Leucine)													
acetolactate synthase I/II/III large subunit [EC:2.2.1.6]	K01652	-	2	1	1	1	1	1	1	1	1	1	-
acetolactate synthase I/II/III small subunit [EC:2.2.1.6]	K01653	-	1	1	1	1	1	1	1	1	1	1	-
ketol-acid reductoisomerase [EC:1.1.1.86]	K00053	-	1	1	1	1	1	1	1	1	1	1	-
dihydroxy-acid dehydratase [EC:4.2.1.9]	K01687	-	1	1	1	1	1	1	1	1	1	1	1
branched-chain amino acid aminotransferase [EC:2.6.1.42]	K00826	-	1	1	1	1	1	1	1	1	1	1	1
alanine-synthesizing transaminase [EC:2.6.1.66 2.6.1.2]	K14260	-	-	-	-	-	-	-	-	-	-	1	1

Table S5.3: Continued.

Functional description	Function ID	Symbol	Fibrobacteria							Chitinivibronia		
			<i>F. succinogenes</i> S85	Fibrobacteraceae			AD_312	Fibromonadaceae			<i>C. alkaliphilus</i> AChtl	MC_77
				AD_80	AD_111	SR_36		IN01_31	IN01_221	IN01_307		
Valine, Isoleucine, leucine biosynthesis (Pyruvate -> Valine/Isoleucine/Leucine)												
2-isopropylmalate synthase [EC:2.3.3.13]	K01649	-	2	2	2	2	2	2	-	2	-	
3-isopropylmalate/(R)-2-methylmalate dehydratase large subunit [EC:4.2.1.33 4.2.1.35]	K01703	-	1	1	1	-	1	1	1	1	2	-
3-isopropylmalate/(R)-2-methylmalate dehydratase small subunit [EC:4.2.1.33 4.2.1.35]	K01704	-	1	1	1	-	1	-	1	1	2	-
3-isopropylmalate dehydrogenase [EC:1.1.1.85]	K00052	-	1	1	1	1	1	1	1	1	1	-
Arginine and Proline metabolism (Glutamate->Arginine)												
Ornithine biosynthesis (glutamate -> ornithine)												
glutamate N-acetyltransferase / amino-acid N-acetyltransferase [EC:2.3.1.35 2.3.1.1]	K00620	-	1	1	1	1	1	1	1	1	1	-
acetylglutamate kinase [EC:2.7.2.8]	K00930	-	1	1	1	1	1	1	1	1	1	-
N-acetyl-gamma-glutamyl-phosphate reductase [EC:1.2.1.38]	K00145	-	1	1	1	1	1	1	1	1	1	-
acetylornithine aminotransferase [EC:2.6.1.11]	K00818	-	1	1	1	1	1	-	-	-	-	-
acetylornithine/N-succinyldiaminopimelate aminotransferase [EC:2.6.1.11 2.6.1.17]	K00821	-	-	-	-	-	-	1	1	1	1	-
Urea cycle												
ornithine carbamoyltransferase [EC:2.1.3.3]	K00611	-	1	1	1	1	1	1	-	1	1	1
arginosuccinate synthase [EC:6.3.4.5]	K01940	-	1	1	1	1	1	1	1	1	1	1
arginosuccinate lyase [EC:4.3.2.1]	K01755	-	1	1	1	1	1	1	-	1	1	1
Proline biosynthesis (Glutamate->Proline)												
glutamate 5-kinase [EC:2.7.2.11]	K00931	-	1	1	1	1	1	1	1	1	1	1
glutamate-5-semialdehyde dehydrogenase [EC:1.2.1.41]	K00147	-	1	1	1	1	1	1	1	1	1	1
pyrroline-5-carboxylate reductase [EC:1.5.1.2]	K00286	-	1	1	1	1	1	1	1	1	1	1
Phenylalanine, Tyrosine, Tryptophan biosynthesis												
Chorismate biosynthesis												
3-deoxy-7-phosphoheptulonate synthase [EC:2.5.1.54]	K01626	-	-	-	-	-	-	-	-	-	2	1
3-dehydroquinate synthase [EC:4.2.3.4]	K01735	-	-	-	-	-	-	-	1	-	1	-
3-dehydroquinate dehydratase I [EC:4.2.1.10]	K03785	-	1	1	1	1	1	1	1	1	-	-
3-dehydroquinate dehydratase II [EC:4.2.1.10]	K03786	-	-	-	-	-	-	-	-	-	1	-
shikimate dehydrogenase [EC:1.1.1.25]	K00014	-	1	1	1	1	1	-	1	1	1	-
shikimate kinase [EC:2.7.1.71]	K00891	-	-	-	-	-	-	1	-	-	2	-
shikimate kinase / 3-dehydroquinate synthase [EC:2.7.1.71 4.2.3.4]	K13829	-	1	1	1	1	1	-	1	1	-	-
3-phosphoshikimate 1-carboxyvinyltransferase [EC:2.5.1.19]	K00800	-	1	1	1	1	1	2	1	2	1	-
chorismate synthase [EC:4.2.3.5]	K01736	-	1	1	1	1	1	1	1	1	1	-
Phenylalanine and Tyrosine biosynthesis (Chorismate -> Phenylalanine/Tyrosine)												
prephenate dehydrogenase [EC:1.3.1.12]	K04517	-	1	1	1	1	1	1	1	1	1	1
chorismate mutase [EC:5.4.99.5]	K01850	-	-	-	-	-	-	-	-	-	1	1
chorismate mutase [EC:5.4.99.5]	K04093	-	1	1	1	1	1	1	1	1	-	-
prephenate dehydratase [EC:4.2.1.51]	K04518	-	1	1	1	1	1	1	1	1	1	1
aspartate aminotransferase [EC:2.6.1.1]	K00812	-	1	1	1	1	1	1	1	1	1	-
aspartate aminotransferase [EC:2.6.1.1]	K11358	-	-	-	-	-	-	-	-	-	1	-
histidinol-phosphate aminotransferase [EC:2.6.1.9]		-	2	1	3	2	2	1	1	1	1	1
Tryptophan biosynthesis (Chorismate -> Tryptophan)												
anthranilate synthase component I [EC:4.1.3.27]	K01657	-	1	1	1	1	1	-	-	1	1	-
anthranilate synthase component II [EC:4.1.3.27]	K01658	-	-	-	-	-	-	1	1	1	1	-

Table 5.3: Continued.

Functional description	Function ID	Symbol	Fibrobacteria							Chitinivibronia		
			Fibrobacteraceae				Fibromonadaceae			<i>C. alkaliphilus</i> ACh1	MC_77	
			<i>F. succinogenes</i> S85	AD_80	AD_111	SR_36	AD_312	IN01_31	IN01_221			IN01_307
Phenylalanine, Tyrosine, Tryptophan biosynthesis												
<i>Tryptophan biosynthesis (Chorismate -> Tryptophan)</i>												
anthranilate synthase/phosphoribosyltransferase [EC:4.1.3.27 2.4.2.18]	K13497	-	1	1	1	1	1	-	-	-	-	-
anthranilate phosphoribosyltransferase [EC:2.4.2.18]	K00766	-	-	-	-	-	-	1	1	1	1	1
phosphoribosylanthranilate isomerase [EC:5.3.1.24]	K01817	-	-	-	-	-	-	1	1	1	1	1
indole-3-glycerol phosphate synthase / phosphoribosylanthranilate isomerase [EC:4.1.1.48 5.3.1.24]	K13498	-	1	1	1	1	1	-	-	-	-	-
indole-3-glycerol phosphate synthase [EC:4.1.1.48]	K01609	-	-	-	-	-	-	1	1	1	1	-
tryptophan synthase alpha chain [EC:4.2.1.20]	K01695	-	1	1	1	1	1	1	1	-	1	1
tryptophan synthase beta chain [EC:4.2.1.20]	K01696	-	1	1	1	1	1	1	1	1	1	-
tryptophan synthase beta chain [EC:4.2.1.20]	K06001	-	1	1	1	1	1	1	-	-	1	-
Cysteine and Methionine biosynthesis												
<i>Cysteine (Serine -> Cysteine)</i>												
serine O-acetyltransferase [EC:2.3.1.30]	K00640	-	2	2	1	2	1	2	2	3	1	-
cysteine synthase A [EC:2.5.1.47]	K01738	-	3	1	2	2	2	2	2	2	3	1
cysteine synthase B [EC:2.5.1.47]	K12339	-	-	-	-	-	-	-	1	1	-	-
<i>Pyruvate -> Cysteine</i>												
cystathione beta-lyase [EC:4.4.1.8]	K14155	-	1	1	-	1	-	1	1	1	-	1
<i>Methionine (aspartate -> homoserine -> methionine)</i>												
aspartate kinase [EC:2.7.2.4]	K00928	-	1	1	1	1	1	1	1	1	1	1
aspartate-semialdehyde dehydrogenase [EC:1.2.1.11]	K00133	-	1	1	1	1	1	1	1	1	1	2
homoserine dehydrogenase [EC:1.1.1.13]	K00003	-	1	1	1	1	1	1	1	1	1	1
homoserine O-succinyltransferase [EC:2.3.1.46]	K00651	-	-	-	-	-	-	-	-	-	1	2
cystathionine gamma-synthase [EC:2.5.1.48]	K01739	-	-	-	-	-	-	1	1	1	1	1
cystathionine beta-lyase [EC:4.4.1.8]	K01760	-	1	-	1	-	-	-	-	-	1	-
5-methyltetrahydrofolate--homocysteine methyltransferase [EC:2.1.1.13]	K00548	-	1	1	1	2	2	2	2	2	1	-
5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase [EC:2.1.1.14]	K00549	-	1	1	1	1	1	-	-	-	1	-
<i>Methionine degradation (Methionine -> Homocysteine->Homoserine)</i>												
S-adenosylmethionine synthetase [EC:2.5.1.6]	K00789	-	1	-	1	1	1	1	1	1	-	-
DNA (cytosine-5)-methyltransferase 1 [EC:2.1.1.37]	K00558	-	1	7	3	3	-	1	4	2	1	1
adenosylhomocysteinase [EC:3.3.1.1]	K01251	-	1	1	1	1	1	-	-	1	-	-
O-acetylhomoserine (thiol)-lyase [EC:2.5.1.49]	K01740	-	1	-	-	-	-	-	-	-	-	-
homoserine O-acetyltransferase [EC:2.3.1.31]	K00641	-	4	3	3	3	1	1	-	1	1	-
<i>Methionine salvage pathway</i>												
S-adenosylmethionine decarboxylase [EC:4.1.1.50]	K01611	-	-	-	-	-	-	-	-	-	1	-
spermidine synthase [EC:2.5.1.16]	K00797	-	1	2	2	1	1	1	1	1	1	1
adenosylhomocysteine nucleosidase [EC:3.2.2.9]	K01243	-	-	-	-	-	-	1	-	-	-	1
5'-methylthioadenosine phosphorylase [EC:2.4.2.28]	K00772	-	-	-	-	-	-	-	1	1	1	-
methylthioribose-1-phosphate isomerase [EC:5.3.1.23]	K08963	-	2	3	2	2	2	1	1	1	1	1
Alanine, Aspartate and Glutamate Metabolism												
<i>Aspartate metabolism</i>												
<i>(L-Aspartate->N-Carbamoyl-L-aspartate/D-Aspartate/Oxaloacetate)</i>												
aspartate carbamoyltransferase catalytic subunit [EC:2.1.3.2]	K00609	-	1	1	1	1	1	-	1	1	1	1
aspartate racemase [EC:5.1.1.13]	K01779	-	1	-	-	-	-	-	-	-	-	-
L-aspartate oxidase [EC:1.4.3.16]	K00278	-	1	1	1	1	1	1	1	1	-	1

Table 5.3: Continued.

Functional description	Function ID	Symbol	Fibrobacteria							Chitinivibronia	
			Fibrobacteraceae				Fibromonadaceae			<i>C. alkaliphilus</i> ACh1	MC_77
			<i>F. succinogenes</i> S85	AD_80	AD_111	SR_36	AD_312	IN01_31	IN01_221		
Alanine, Aspartate and Glutamate Metabolism											
<i>(L-Aspartate->Fumarate)</i>											
argininosuccinate synthase [EC:6.3.4.5]	K01940	-	1	1	1	1	1	1	1	1	1
argininosuccinate lyase [EC:4.3.2.1]	K01755	-	1	1	1	1	1	1	-	1	1
adenylosuccinate synthase [EC:6.3.4.4]	K01939	-	1	1	1	1	1	1	1	2	1
adenylosuccinate lyase [EC:4.3.2.2]	K01756	-	1	1	1	1	1	1	1	1	1
aspartate ammonia-lyase [EC:4.3.1.1]	K01744	-	-	-	-	-	-	-	-	-	1
<i>Aspartate synthesis</i>											
<i>(N-Acetyl-L-aspartate->L-Aspartate)</i>											
aspartoacylase [EC:3.5.1.15]	K01437	-	1	1	1	1	1	1	1	1	-
<i>(L-Asparagine->L-Aspartate)</i>											
L-asparaginase [EC:3.5.1.1]	K01424	-	1	1	1	1	-	-	-	-	-
<i>(Oxaloacetate<->L-Aspartate)</i>											
aspartate aminotransferase [EC:2.6.1.1]	K00812	-	1	1	1	1	1	1	1	1	1
aspartate aminotransferase [EC:2.6.1.1]	K11358	-	-	-	-	-	-	-	-	-	1
<i>Glutamate synthesis and metabolism</i>											
<i>(2-Oxo-glutarate->Glutamate)</i>											
glutamate dehydrogenase [EC:1.4.1.2]	K00260	-	-	-	-	-	-	-	-	-	1
glutamate dehydrogenase (NADP+) [EC:1.4.1.4]	K00262	-	1	1	1	1	-	1	1	1	1
glutamate synthase (NADPH/NADH) large chain [EC:1.4.1.13 1.4.1.14]	K00265	-	1	1	1	1	-	-	-	-	-
glutamate synthase (NADPH/NADH) small chain [EC:1.4.1.13 1.4.1.14]	K00266	-	2	2	2	2	1	1	1	1	1
<i>(Glutamate->Glutamine)</i>											
glutamine synthetase [EC:6.3.1.2]	K01915	-	2	2	2	3	2	2	1	2	2
<i>(Glutamine->5-Phosphoribosylamine/Carbomoylphosphate)</i>											
amidophosphoribosyltransferase [EC:2.4.2.14]	K00764	-	2	2	2	2	1	1	1	1	1
carbamoyl-phosphate synthase large subunit [EC:6.3.5.5]	K01955	-	4	1	1	1	2	1	1	1	-
carbamoyl-phosphate synthase small subunit [EC:6.3.5.5]	K01956	-	1	1	1	1	1	1	1	1	1
<i>(D-Glucosamine-6P->Glutamine)</i>											
glucosamine--fructose-6-phosphate aminotransferase (isomerizing) [EC:2.6.1.16]	K00820	-	1	1	1	1	1	1	1	1	1
<i>Alanine synthesis and metabolism</i>											
<i>(Alanine<->Pyruvate)</i>											
alanine-synthesizing transaminase [EC:2.6.1.66 2.6.1.2]	K14260	-	-	-	-	-	-	-	-	-	1
<i>(Asparagine->Alanine)</i>											
L-asparaginase [EC:3.5.1.1]	K01424	-	1	1	1	1	-	-	-	-	-
Lysine biosynthesis (aspartate->lysine)											
aspartate kinase [EC:2.7.2.4]	K00928	-	1	1	1	1	1	1	1	1	1
aspartate-semialdehyde dehydrogenase [EC:1.2.1.11]	K00133	-	1	1	1	1	1	1	1	1	2
4-hydroxy-tetrahydrodipicolinate synthase [EC:4.3.3.7]	K01714	-	1	1	1	1	1	1	1	1	2
4-hydroxy-tetrahydrodipicolinate reductase [EC:1.17.1.8]	K00215	-	1	1	1	1	1	1	1	1	2
diaminopimelate dehydrogenase [EC:1.4.1.16]	K03340	-	1	1	2	1	1	-	-	-	-
Histidine metabolism (PPP -> Histidine)											
ATP phosphoribosyltransferase [EC:2.4.2.17]	K00765	-	1	1	1	1	1	-	1	-	1
phosphoribosyl-ATP pyrophosphohydrolase [EC:3.6.1.31]	K01523	-	1	1	1	1	1	1	1	1	-
phosphoribosyl-AMP cyclohydrolase [EC:3.5.4.19]	K01496	-	1	1	1	1	1	1	1	1	1
phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase [EC:5.3.1.16]	K01814	-	1	1	1	1	1	1	1	1	2

Table 5.3: Continued.

Functional description	Function ID	Symbol	Fibrobacteria							Chitinivibronia		
			Fibrobacteraceae				Fibromonadaceae			<i>C. alkaliphilus</i> ACh1	MC_77	
			<i>F. succinogenes</i> S85	AD_80	AD_111	SR_36	AD_312	IN01_31	IN01_221			IN01_307
Histidine metabolism (PPP -> Histidine)												
glutamine amidotransferase [EC:2.4.2.-]	K02501	-	1	1	1	1	1	1	1	1	1	1
cyclase [EC:4.1.3.-]	K02500	-	1	1	1	1	1	1	1	1	1	1
imidazoleglycerol-phosphate dehydratase [EC:4.2.1.19]	K01693	-	1	1	1	1	1	1	2	-	1	1
histidinol-phosphate aminotransferase [EC:2.6.1.9]	K00817	-	2	1	3	2	2	1	1	1	1	1
histidinol-phosphatase (PHP family) [EC:3.1.3.15 4.2.1.19]	EC 3.1.3.15 EC 4.2.1.19	-	-	-	-	-	-	-	-	-	-	-
histidinol dehydrogenase [EC:1.1.1.23]	K00013	-	1	1	1	1	1	1	2	-	1	1
Metabolism of cofactors and vitamins												
Thiamine metabolism												
thiamine-phosphate pyrophosphorylase [EC:2.5.1.3]	K00788	-										
hydroxymethylpyrimidine/phosphomethylpyrimidine kinase [EC:2.7.1]	K00941	-	1	1	2	1	2	2	2	2	3	2
thiamine-monophosphate kinase [EC:2.7.4.16]	K00946	-	1	1	1	1	1	1	1	1	0	0
1-deoxy-D-xylulose-5-phosphate synthase [EC:2.2.1.7]	K01662	-	1	1	1	1	1	1	1	1	1	0
phosphomethylpyrimidine synthase [EC:4.1.99.17]	K03147	-	2	2	2	2	1	1	1	1	1	1
sulfur carrier protein ThiS adenylyltransferase [EC:2.7.7.73]	K03148	-	1	1	0	1	1	1	1	1	1	0
thiazole synthase [EC:2.8.1.10]	K03149	-	1	1	1	1	0	0	0	0	1	0
2-iminoacetate synthase [EC:4.1.99.19]	K03150	-	1	1	1	1	1	1	1	1	1	1
thiamine biosynthesis protein ThiI	K03151	-	1	1	1	1	1	2	2	2	2	2
cysteine desulfurase [EC:2.8.1.7]	K04487	-	0	0	0	0	0	0	0	0	1	0
Riboflavin metabolism												
riboflavin synthase [EC:2.5.1.9]	K00793	-	1	1	1	0	0	0	0	0	3	0
6,7-dimethyl-8-ribityllumazine synthase [EC:2.5.1.78]	K00794	-	1	1	1	1	1	1	1	1	1	0
diaminohydroxyphosphoribosylaminopyrimidine deaminase / 5-amino-6-(5-phosphoribosylamino)uracil reductase [EC:3.5.4.26 1.1.1.193]	K11752	-	1	0	0	1	1	0	1	1	1	0
riboflavin kinase / FMN adenylyltransferase [EC:2.7.1.26 2.7.7.2]	K11753	-	1	1	1	1	1	1	1	1	1	1
3,4-dihydroxy 2-butanone 4-phosphate synthase / GTP cyclohydrolase II [EC:4.1.99.12 3.5.4.25]	K14652	-	1	1	1	1	1	1	1	1	1	1
Vitamin B6 metabolism												
4-hydroxythreonine-4-phosphatedehydrogenase[EC:1.1.1.262]	K00097	-	1	0	0	1	1	1	1	1	1	0
phosphoserineaminotransferase[EC:2.6.1.52]	K00831	-	0	0	0	0	1	1	1	1	1	1
pyridoxinekinase[EC:2.7.1.35]	K00868	-	1	1	0	1	1	1	1	1	1	1
threoninesynthase[EC:4.2.3.1]	K01733	-	1	0	1	1	0	0	0	0	0	0
erythronate-4-phosphatedehydrogenase[EC:1.1.1.290]	K03473	-	1	1	1	1	1	1	1	1	1	1
pyridoxine5'-phosphatesynthase[EC:2.6.99.2]	K03474	-	0	0	0	0	1	1	1	1	1	1
pyridoxal5'-phosphatesynthasepdxSsubunit[EC:4.3.3.6]	K06215	-	1	1	1	1	1	1	1	1	1	1
5'-phosphatesynthasepdxTsubunit[EC:4.3.3.6]	K08681	-	1	1	1	1	0	0	0	0	0	0
Nicotinate and nicotinamide metabolism												
L-aspartate oxidase [EC:1.4.3.16]	K00278	-	1	1	1	1	1	1	1	1	1	0
quinolinate synthase [EC:2.5.1.72]	K03517	-	1	1	1	1	1	1	1	1	1	1
nicotinate-nucleotide pyrophosphorylase (carboxylating) [EC:2.4.2.19]	K00767	-	1	1	1	1	1	1	1	1	1	1
5'-nucleotidase [EC:3.1.3.5]	K03787	-	1	1	1	1	1	1	1	1	1	1
purine-nucleoside phosphorylase [EC:2.4.2.1]	K03783	-	0	0	0	0	0	0	0	0	1	0
NAD+ diphosphatase [EC:3.6.1.22]	K03426	-	1	1	1	1	0	1	1	0	1	0
nicotinate-nucleotide adenylyltransferase [EC:2.7.7.18]	K00969	-	1	1	1	1	1	0	0	0	1	0

Table 5.3: Continued.

Functional description	Function ID	Symbol	Fibrobacteria							Chitinivibronia		
			Fibrobacteraceae				Fibromonadaceae			<i>C. alkaliphilus</i> AChtl	MC_77	
			<i>F. succinogenes</i> S85	AD_80	AD_111	SR_36	AD_312	IN01_31	IN01_221			IN01_307
Nicotinate and nicotinamide metabolism												
NAD ⁺ synthase (glutamine-hydrolysing) [EC:6.3.5.1]	K01950	-	2	1	1	2	1	0	0	0	1	0
NAD ⁺ synthase [EC:6.3.1.5]	K01916	-	0	0	0	0	0	1	1	1	0	0
5'-nucleotidase [EC:3.1.3.5]	K01081	-	0	0	0	0	0	1	0	0	0	0
NAD ⁺ kinase [EC:2.7.1.23]	K00858	-	1	1	1	1	1	1	1	1	1	1
NAD(P) transhydrogenase subunit alpha [EC:1.6.1.2]	K00324	-	0	0	0	0	0	0	0	0	2	0
NAD(P) transhydrogenase subunit beta [EC:1.6.1.2]	K00325	-	0	0	0	0	0	0	0	0	1	0
nicotinamide-nucleotide amidase [EC:3.5.1.42]	K03742	-	0	0	0	0	1	0	0	0	0	0
nicotinamide-nucleotide amidase [EC:3.5.1.42]	K03743	-	1	1	1	1	0	0	0	0	1	1
bifunctional NMN adenylyltransferase/nudix hydrolase [EC:2.7.7.1 3.6.1.-]	K13522	-	0	0	0	0	0	1	1	0	0	0
Pantothenate and CoA biosynthesis												
keto-acid reductoisomerase [EC:1.1.1.86]	K00053	-	1	1	1	1	1	1	1	1	1	0
2-dehydropantoate 2-reductase [EC:1.1.1.169]	K00077	-	1	0	0	0	0	0	0	0	0	0
3-methyl-2-oxobutanoate hydroxymethyltransferase [EC:2.1.2.11]	K00606	-	1	1	1	1	1	1	1	1	1	1
branched-chain amino acid aminotransferase [EC:2.6.1.42]	K00826	-	1	1	1	1	1	1	1	1	1	1
dephospho-CoA kinase [EC:2.7.1.24]	K00859	-	1	1	1	1	1	0	0	0	1	1
pantetheine-phosphate adenylyltransferase [EC:2.7.7.3]	K00954	-	1	1	1	1	1	0	2	1	1	0
holo-[acyl-carrier protein] synthase [EC:2.7.8.7]	K00997	-	0	0	0	0	0	0	0	0	1	1
aspartate 1-decarboxylase [EC:4.1.1.11]	K01579	-	1	1	1	1	1	1	1	1	1	1
acetolactate synthase I/II/III large subunit [EC:2.2.1.6]	K01652	-	2	1	1	1	1	1	1	1	1	0
acetolactate synthase I/II small subunit [EC:2.2.1.6]	K01653	-	1	1	0	1	1	1	1	1	1	0
dihydroxy-acid dehydratase [EC:4.2.1.9]	K01687	-	1	1	1	1	1	1	1	1	1	1
pantoate--beta-alanine ligase [EC:6.3.2.1]	K01918	-	1	1	1	1	1	1	1	1	1	1
type III pantothenate kinase [EC:2.7.1.33]	K03525	-	1	0	2	1	1	1	1	1	0	0
4'-phosphopantetheinyl transferase [EC:2.7.8.-]	K06133	-	1	0	0	1	0	0	0	0	0	0
phosphopantothenoylcysteine decarboxylase / phosphopantothenate- -cysteine ligase [EC:4.1.1.36 6.3.2.5]	K13038	-	1	1	1	1	1	0	1	1	1	1
Biotin metabolism												
3-oxoacyl-[acyl-carrier protein] reductase [EC:1.1.1.100]	K00059	-	4	3	2	2	2	2	1	2	2	1
enoyl-[acyl-carrier protein] reductase I [EC:1.3.1.9 1.3.1.10]	K00208	-	0	0	0	0	0	1	1	0	1	0
3-oxoacyl-[acyl-carrier-protein] synthase I [EC:2.3.1.41]	K00647	-	1	0	0	0	0	1	0	0	0	0
8-amino-7-oxononanoate synthase [EC:2.3.1.47]	K00652	-	0	0	0	0	1	1	1	1	1	0
adenosylmethionine-8-amino-7-oxononanoate aminotransferase [EC:2.6.1.62]	K00833	-	1	1	1	1	1	0	1	1	1	0
biotin synthase [EC:2.8.1.6]	K01012	-	2	2	2	2	1	2	2	2	2	2
6-carboxyhexanoate--CoA ligase [EC:6.2.1.14]	K01906	-	1	1	1	1	0	0	0	0	0	0
dethiobiotin synthetase [EC:6.3.3.3]	K01935	-	1	1	1	1	1	1	1	1	1	0
malonyl-CoA O-methyltransferase [EC:2.1.1.197]	K02169	-	0	0	0	0	1	1	1	1	0	1
3-hydroxyacyl-[acyl-carrier-protein] dehydratase [EC:4.2.1.59]	K02372	-	2	2	2	2	1	1	1	1	0	0
BirA family transcriptional regulator, biotin operon repressor / biotin- [acetyl-CoA-carboxylase] ligase [EC:6.3.4.15]	K03524	-	1	1	1	1	1	1	1	1	1	0
3-oxoacyl-[acyl-carrier-protein] synthase II [EC:2.3.1.179]	K09458	-	4	3	3	3	3	2	1	1	3	1
biotin synthesis protein BioG	K09789	-	0	0	0	0	1	0	0	0	0	0
Lipoic acid metabolism												
lipoic acid synthetase [EC:2.8.1.8]	K03644	-	0	0	0	0	0	0	0	0	1	1
lipoate-protein ligase A [EC:2.7.7.63]	K03800	-	0	0	0	0	0	0	0	0	1	2
Folate biosynthesis												
dihydrofolatereductase [EC:1.5.1.3]	K00287	-	1	1	1	1	1	1	1	1	0	0

Table 5.3: Continued.

Functional description	Function ID	Symbol	Fibrobacteria							Chitinivibronia		
			Fibrobacteraceae				Fibromonadaceae			<i>C. alkaliphilus</i> AChtl	MC_77	
			<i>F. succinogenes</i> S85	AD_80	AD_111	SR_36	AD_312	IN01_31	IN01_221			IN01_307
Folate biosynthesis												
dihydropteratesynthase[EC:2.5.1.15]	K00796	-	1	1	1	1	1	1	1	1	1	1
2-amino-4-hydroxy-6-hydroxymethyl-dihydropteridinediphosphokinase[EC:2.7.6.3]	K00950	-	1	1	1	1	1	0	0	1	1	1
alkalinephosphatase[EC:3.1.3.1]	K01077	-	0	0	0	0	1	0	0	0	0	0
GTPcyclohydrolase[EC:3.5.4.16]	K01495	-	1	1	1	2	1	1	1	1	1	1
dihydroneopterinaldolase[EC:4.1.2.25]	K01633	-	1	1	1	1	1	1	1	1	0	0
para-aminobenzoatesynthetasecomponent[EC:2.6.1.85]	K01665	-	0	0	0	0	0	0	0	0	1	0
6-pyruvoyltetrahydropterin-6-carboxytetrahydropterinsynthase[EC:4.2.3.124.1.2.50]	K01737	-	1	2	0	2	1	1	0	0	2	1
4-amino-4-deoxychorismatylase[EC:4.1.3.38]	K02619	-	0	0	0	0	0	0	0	0	1	0
para-aminobenzoatesynthetase/4-amino-4-deoxychorismatylase [EC:2.6.1.85.4.1.3.38]	K03342	-	1	1	0	0	0	0	0	0	0	0
7-cyano-7-deazaguaninesynthase[EC:6.3.4.20]	K06920	-	1	2	0	2	0	0	0	0	1	0
7-cyano-7-deazaguaninereductase[EC:1.7.1.13]	K09457	-	1	1	1	1	0	1	1	1	1	1
7-carboxy-7-deazaguaninesynthase[EC:4.3.99.3]	K10026	-	1	1	0	1	0	0	0	0	1	1
dihydrofolatesynthase/folypolyglutamatesynthase [EC:6.3.2.12.6.3.2.17]	K11754	-	1	1	1	1	1	1	1	1	1	0
One carbon pool by folate												
dihydrofolate reductase [EC:1.5.1.3]	K00287	-	1	1	1	1	1	1	1	1	0	0
methylenetetrahydrofolate reductase (NADPH) [EC:1.5.1.20]	K00297	-	1	1	1	1	1	1	1	1	1	0
5-methyltetrahydrofolate--homocysteine methyltransferase [EC:2.1.1.13]	K00548	-	1	1	1	1	1	1	1	1	1	0
thymidylate synthase [EC:2.1.1.45]	K00560	-	1	1	1	1	1	1	1	0	0	1
glycine hydroxymethyltransferase [EC:2.1.2.1]	K00600	-	1	1	1	1	1	1	1	1	1	0
phosphoribosylaminoimidazolecarboxamide formyltransferase / IMP cyclohydrolase [EC:2.1.2.3.3.5.4.10]	K00602	-	1	0	1	1	1	1	1	1	1	1
methionyl-tRNA formyltransferase [EC:2.1.2.9]	K00604	-	1	1	1	1	1	1	1	1	1	0
aminomethyltransferase [EC:2.1.2.10]	K00605	-	0	0	0	0	0	0	0	0	0	1
formyltetrahydrofolate deformylase [EC:3.5.1.10]	K01433	-	1	1	1	1	1	1	0	0	1	0
methylenetetrahydrofolate dehydrogenase (NADP+)												
/methylenetetrahydrofolate cyclohydrolase [EC:1.5.1.5.3.5.4.9]	K01491	-	1	1	1	1	1	1	1	1	1	0
5-formyltetrahydrofolate cyclo-ligase [EC:6.3.3.2]	K01934	-	1	1	1	1	1	0	0	0	1	1
phosphoribosylglycinamide formyltransferase 2 [EC:2.1.2.2]	K08289	-	1	1	1	1	1	1	1	0	1	0
phosphoribosylglycinamide formyltransferase 1 [EC:2.1.2.2]	K11175	-	1	1	1	1	1	1	1	1	1	0
Ubiquinone and other terpenoid-quinone biosynthesis												
1,4-dihydroxy-2-naphthoate octaprenyltransferase [EC:2.5.1.74.2.5.1.-]	K02548	-	0	0	0	0	0	0	0	0	1	0
4-hydroxybenzoate octaprenyltransferase [EC:2.5.1.-]	K03179	-	2	1	2	2	2	1	1	1	0	0
3-octaprenyl-4-hydroxybenzoate carboxy-lyase UbiD [EC:4.1.1.-]	K03182	-	1	1	1	1	1	1	1	1	0	0
ubiquinone/menaquinone biosynthesis methyltransferase [EC:2.1.1.163.2.1.1.201]	K03183	-	1	1	1	1	1	1	1	1	0	0
3-octaprenyl-4-hydroxybenzoate carboxy-lyase UbiX [EC:4.1.1.-]	K03186	-	1	1	1	1	1	1	1	1	0	0
chorismate dehydratase [EC:4.2.1.151]	K11782	-	1	1	1	1	1	1	1	1	0	0
fitalosine hydrolase [EC:3.2.2.26]	K11783	-	1	1	1	1	1	1	1	0	0	0
cyclic dehydropoxanthinyl fitalosine synthase [EC:1.21.99.2]	K11784	-	0	0	0	0	1	0	0	0	0	0
1,4-dihydroxy-6-naphthoate synthase [EC:1.14.-.-]	K11785	-	1	1	1	1	1	1	1	1	0	0
aminodeoxyfitalosine synthase [EC:2.5.1.120]	K18285	-	1	1	1	1	1	1	0	0	0	0