THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Genomic- and culture-enabled analyses of host-associated methanogenic archaea

James Gregory Volmer

Bachelor of Science with Majors in Biochemistry and Molecular Biology and Microbiology

Bachelor of Science with Honours Class I in the Field of Microbiology

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2021

Faculty of Medicine

Abstract

Methanogenic archaea (methanogens) represent a unique group of organisms that produce methane as a by-product of their anaerobic energy production pathway known as methanogenesis. These methanogens play an important role by maintaining favourable conditions for bacterial fermentation in both host-associated and environmental communities. With implementation of next-generation sequencing technologies and the subsequent identification of complex archaeal communities, there has been an increasing interest the larger role of methanogens - from the targeted reduction of agricultural methane emissions to the implication of specific methanogen species in human health and disease. With this in mind, the key aims of my PhD research are the following: 1) to characterise the effect of a disease-like gut ecology on the community and expression of methanogens; 2) to utilise available metagenomic data to expand the identification and characterisation of methanogen species, and their association with gastrointestinal health and disease; 3) to culture novel methanogen representatives from different host species to expand our understanding of unique archaeal communities. This work initially focused on the human gastrointestinal (GIT) and associated archaea but later expanded to include non-human hosts as a part of a collaboration with the Australian Centre of Ecogenomics and funding by Meat and Livestock Australia. Here, further characterisation of the methanogen communities from non-human hosts is conducted, including the first isolation of novel methanogen lineages from several marsupial species.

Chapter two describes the effects of bile salt, a key factor in many GIT diseases and disorders, on human *Methanobrevibacter* and *Methanosphaera*. New strains of human *M. smithii* (JC9 and PAM) and *M. stadtmanae* (PA5) were successfully isolated, with the latter representing only the second available isolate for this species. Comparative analysis displayed differential bile tolerance between the strains, particularly with *M. stadtmanae* for which the type strain was significantly affected by increasing bile salt concentration and isolate PA5 was resistant. Subsequently, metabolomic analysis of these culture supernatants showed *M. stadtmanae* DSMZ3091^T to produce a unique bile acid metabolite profile, with *M. stadtmanae* PA5 and the *M. smithii* strains all clustering separately. Potentially explaining this differentiation, *M. stadtmanae* PA5 shows a unique bile salt hydrolase gene structure and a greater fold increase in expression in response to bile salt. Together this shows a differential response of human methanogens to bile salt and could affect their response to altered gut ecology.

Chapter three focuses on the expansion of the human associated methanogens through the recovery of metagenome-assembled genomes (MAGs) and subsequent comparative analyses. Along with available methanogen genomes, recovery and analysis of human MAGs significantly expands the

number of species within the methanogen community, with the recovery of novel *Methanosphaera, Methanobacterium*, and multiple *Methanomassiliicoccales*-associated species. Phylogenetic analysis of *M. smithii* shows two distinct clades: *M. smithii* and *M. smithii*_A. Comparative genomic analyses between these clades shows a differential enrichment in metal uptake, carbon assimilation, phosphonate metabolism and glycosylation-associated genes. Additionally, I also show a potential selection for *M. smithii*_A in patients with Crohn's disease compared to ulcerative colitis, which produced MAGs from both lineages. These differentiations in *M. smithii* phylogroups, and the respective genomic potential, may be integral to understanding the community shifts observed in different disease states.

Chapter four works to characterise host-associated *Methanocorpusculum* as a dominant methanogen species in specific animal hosts and to define the association genetic potential. The methanogen community of different marsupial species was characterised, with *Methanocorpusculum* identified as a dominant lineage. Subsequently, *Methanocorpusculum vombatium* and *Methanocorpusculum petaurusium* were isolated from the faecal samples of a common wombat and mahogany glider, respectively, and represent the first host-associated *Methanocorpusculum* to be isolated. Representatives of *Methanomethylophilaceae* and *Methanobrevibacter* were also recovered from mahogany glider and kangaroo samples, respectively. *Methanocorpusculum* MAGs were successfully recovered from 13 animal species, providing a significant expansion of the *Methanocorpusculum* genus. Comparison within and between clades show differential enrichment environmental and host-associated metal transport, amino acid biosynthesis, virulence, and glycosylation proteins, likely providing their respective host specificity. This work significantly expands the identification and characterisation of novel host-associated *Methanocorpusculum*, and identifies their unique genomic potential compared to environmental isolates.

Chapter five provides a general discussion and summary of my PhD research, which provides a significant expansion of available methanogen species from human and non-human animal hosts and characterises their genomic and functional potential. I identify key genetic factors which are differentially enriched between different *Methanobrevibacter* and *Methanosphaera*, with an application toward different functional properties. I also provide a significant expansion of the genus *Methanocorpusculum*, with the comprehensive analysis and identification of host-associated clades supported by the isolation of the first host-associated species: *M. vombatium* and *M. petaurusium*. Collectively, these novel findings further expand our understanding of methanogenic archaea and the role they play in human and non-human GIT ecology.

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, financial support 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 higher degree by research 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 and have sought permission from co-authors for any jointly authored works included in the thesis.

Publications included in this thesis

No publications included.

Submitted manuscripts included in this thesis

No manuscripts submitted for publication.

Other publications during candidature

Published Peer-revied Papers

Hoedt, E. C., D. H. Parks, **J. G. Volmer**, C. P. Rosewarne, S. E. Denman, C. S. McSweeney, J. G. Muir, P. R. Gibson, P. O. Cuiv, P. Hugenholtz, G. W. Tyson and M. Morrison (2018). "Culture- and metagenomics-enabled analyses of the Methanosphaera genus reveals their monophyletic origin and differentiation according to genome size." Isme j **12**(12): 2942-2953.

Conference Abstracts

Emily C. Hoedt**, James G. Volmer**, Páraic Ó Cuív, and Mark Morrison. "Bile salt hydrolase genes in human Methanobrevibacter and Methanosphaera spp." 26-28th June 2018. International Human Microbiome Consortium Meeting 2018. Killarney, Ireland. Not in Attendance.

James G. Volmer, Emily C. Hoedt, Páraic Ó Cuív, and Mark Morrison. "Bile salt hydrolase genes in human Methanobrevibacter and Methanosphaera spp." 26th July 2018. Translational Research Symposium (TRS). Translational Research Institute, Brisbane, Australia.

James G. Volmer, Emily C. Hoedt, Páraic Ó Cuív, and Mark Morrison. "Bile salt hydrolase genes in human Methanobrevibacter and Methanosphaera spp." 1st August 2018. PAH Health Symposium. Princess Alexandra Hospital, Brisbane, Australia.

James G. Volmer, Emily C. Hoedt, Páraic Ó Cuív, and Mark Morrison. "Bile salt hydrolase genes in human Methanobrevibacter and Methanosphaera spp." 22nd November 2018. Clinical and Public Health Postgraduate Symposium. Royal Brisbane and Women's Hospital, Brisbane, Australia.

James G. Volmer and Mark Morrison. "Comparative genome analyses of human gut methanogens". 15-17th April 2019. 2019 Congress on Gastrointestinal Function. Chicago, USA. Not in Attendance.

James G. Volmer, Ben J. Woodcroft, Gene W. Tyson and Mark Morrison. "Differential Enrichment of Metal Transport Proteins in Specific Phylo-groups of Human Methanobrevibacter". 9th October 2019. Centre for Metals in Biology Symposium. University of Queensland, Brisbane, Australia.

James G. Volmer, Emily C. Hoedt, Páraic Ó Cuív and Mark Morrison. "Methanogenic archaea and their role gastrointestinal health and disease". 4-8th November 2019. Sheep CRC/MLA/APL Postgraduate Conference, Manly, NSW, Australia. Oral presentation.

James G. Volmer. "Human methanogenic archaea and their role gastrointestinal health and disease". 22nd November 2019. Australian Gastrointestinal Research Alliance and Princess Alexandra Hospital Meeting, TRI, Australia. Oral Presentation

James G. Volmer. "Methanogens (microorganisms that produce methane) in humans, what causes them to thrive, their impact on health, and what might be done to reduce their impact – with reference to and comparison with ruminants". 24th May 2021. Oral Presentation

James G. Volmer, Vinod K. Narayana, Dedreia L. Tull, Malcolm J. McConville, and Mark Morrison. "The human methanogens Methanobrevibacter and Methanosphaera have differential effects on the bile acid pool as revealed by metabolomic analysis" International Human Microbiome Consortium Congress 2021. 27-29th June 2021. Virtual Conference.

Contributions by others to the thesis

The work presented in this thesis was critically reviewed by my supervisor Prof. Mark Morrison.

Chapter 2: The metabolomics data was generated by Metabolomics Australia, in collaboration with Vinod K. Narayana, Dedreia L. Tull, and Malcolm J. McConville

Chapter 3: None

Chapter 4: Marsupial sample collection was conducted by DS Teakle and Amy Shima. Marsupial faecal DNA extraction and metagenome sequencing, along with the recovery of archaeal MAGs was conducted by Dr. Rochelle Soo. Marsupial faecal culture gas production data was generated by Dr. Emily Hoedt. Enrichment of *Methanocorpusculum* sp. CW153 was conducted with the assistance of Ana L. Astorga

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

The isolation of methanogen isolates *Methanobrevibacter smithii* JC9 and *Methanosphaera stadtmanae* PA5, as described in Section 2.3.1, were recovered as a part of a BSc Honours at the University of Queensland awarded on the 21st July 2017. All subsequent data generated using these isolates is novel to this thesis.

Research Involving Human or Animal Subjects

Human faecal samples were collected under the UQ Human Research Ethics Approval (HREC) #2015000775. Additional samples were for a nutritional trial performed at Monash University under MU-HREC CF14/2904 – 2014001593 and UQ-HREC #2015000317 (*M. smithii* PAM sample). Ethical permission for the collection of all marsupial faecal samples was granted by the Animal Welfare Unit, the University of Queensland, Brisbane, Australia under ANRFA/SCMB/099/14.

Acknowledgements

To my supervisors, I am extremely grateful for their support throughout the years. A very special thank you to my primary supervisor Prof. Mark Morrison for his guidance and mentorship ever since starting my Honours in 2016, the knowledge and opportunities you provided me with were invaluable and I will carry them forward with me. To Dr. Gene Tyson and Dr. Gerald Holtmann, I am grateful for your advice and support of my research through the project.

I would like to acknowledge the University of Queensland Diamantina Institute for providing world-class facilities and expertise, allowing me to produce the highest quality of research. Additionally, I would like to gratefully acknowledge Meat and Livestock Australia for their support through a top-up scholarship and technical assistance grant, which provided me with the opportunity to expand my research.

I am extremely grateful to everyone at the UQ Diamantina Institute and Translational Research Institute for your support and friendship. I would like to thank all Morrison group members, past and present, for their help with lab work, as well as their friendship and guidance over the years. A special thank you to Jing Jie, it has been a long ride since Honours, but I have appreciated your friendship and all your help along the way. Thank you to Dr. Emily Hoedt for her guidance through my Honours and the start of my Ph.D.

Thank you to Prof. Phillip Hugenholtz and the Australian Centre of Ecogenomics for providing access to their bioinformatic facilities and thank you to Dr. Ben Woodcroft for his guidance and support with the bioinformatic analyses. Thank you to Prof. Chris McSweeney, Dr. Stuart Denman, and Jagadish Padmanabha of CSIRO Agriculture and Food for allowing me to use the gassing facilities, gas chromatography, and fluorescence microscopy facilities. Thank you to Jag for his additional technical assistance and for his help with any questions I had. I would also like to acknowledge Prof. Malcolm McConville, Dr. Dedreia Tull, and Dr. Vinod Narayana of the University of Melbourne Bio21 for the metabolomics analysis and technical support with the analyses.

Lastly, a massive thank you to all my family and friends, I could not have done it without you. Thank you to my parents for their support throughout the last few years, for all the warm meals, and for putting up with the late nights. Thank you to Keanu for the late-night gaming sessions that kept me sane. And thank you to my girlfriend Amy for supporting me through the long hours and for the baked goods to help get me through.

Financial support

This research was supported by an Australian Government Research Training Program Scholarship. This research was also supported from January 1st 2019 to July 1st 2021 by Meat and Livestock Australia under project no. B.STU.1909.

Keywords

Archaea, methanogen, methane, bile salt, Methanobrevibacter, Inflammatory bowel disease, marsupial, Methanocorpusculum,

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 060503, Microbial Genetics, 20%

ANZSRC code: 060599, Microbiology not elsewhere classified, 40%

ANZSRC code: 060309, Phylogeny and Comparative Analysis, 40%

Fields of Research (FoR) Classification

FoR code: 0605, Microbiology, 60%

FoR code: 0604, Genetics, 20%

FoR code: 0603, Evolutionary Biology, 20%

The Effect of COVID-19 on PhD Research Direction

This section will focus on describing the effects of the COVID-19 pandemic on my PhD research.

I spent time in Japan and Korea in February 2020, returning to Australian on March 3rd. Subsequently, I was in self-quarantine for the following two weeks until March 18th. Due to COVID-19, my laboratory-based activities at the Translational Research Institute (TRI) ceased between March 23rd through April 26th (5 weeks). From April 27th through June 21st (8 weeks), my research group was separated into two groups. I could only attend the lab on alternating workdays and stayed home on public holidays (i.e. 4-5/10 days per fortnight). From June 22nd, I was permitted to work daily but with limited hours to comply with the building occupancy restrictions at TRI until mid-July. Due to personal health issues placing me in a higher risk category, I spent additional time away from the lab through June and July, returning to 'normal' lab hours in August.

I estimate these restrictions reduced by lab-based activity by ~16 weeks and affected me in the following way:

1. In total, this reduced lab-based functional analyses and resulted in a larger focus on bioinformatic analyses for my research.
2. Originally, the Chapter three of my PhD research was to characterise the gut methanogen community of individuals in a placebo controlled randomised trial of probiotic intervention to correct functional constipation. Recruitment and sampling were to occur in early 2020 but was ultimately halted due to COVID-19.
3. During this time, my work completely shifted to bioinformatics, which allowed me to expand my analyses to host-associated methanogens from non-human animals. This aim further involved the characterisation and isolation of methanogens from Australian marsupials, in collaboration with the Australian Centre of Ecogenomics. This provided an opportunity to focus on a further expansion and characterisation of the methanogen community from non-human animals.

Table of Contents

List of Figures in Thesis

List of Tables in Thesis

List of Abbreviations

°C – degrees Celsius

µg – microgram

µL – microliter

16S rRNA – 16S ribosomal RNA

AA – auxiliary activity

AAI – Average Amino acid Identity

ACE – Australian Centre for Genomics

acs – acetyl-CoA synthetase

AcuB – acetoin utilization protein B

Adh – Alcohol dehydrogenase

AGRF – Australian Genome Research Facility

alc – alcohol dehydrogenase

ald – aldehyde dehydrogenase

ALP – adhesin-like protein

ANI – average nucleotide identity

ANOVA – One-way Analysis of Variance

ATP – adenosine tri-phosphate

badF – Benzoyl-CoA reductase subunit A

BAL – bronchoalveolar lavage fluid

BLaER1 – Human B-cell Precursor Leukemia Cell Line

BLASTp – basic local alignment search tool- protein

Bp – base pair

BPGA – Bacterial Pan Genome Analysis

BRN-RF10 – Balch 10% rumen fluid medium

BRN-RF30 – Balch 30% rumen fluid medium

Bsh/*bsh* – bile salt hydrolase

C1 – one-carbon

Caco-2/BBe – Caucasian colon adenocarcinoma model cell line

CAZyme – carbohydrate-active enzymes

CBM – carbohydrate-binding module

CD – Crohn's Disease

CDI – *Clostridium difficile* infection

cDNA – complementary DNA

CE – Carbohydrate esterase

CH_4 – methane

CKD – Chronic Kidney Disease

CO_2 – carbon dioxide

COG – clusters of orthologous groups

CoM – Coenzyme M

CRC – colorectal cancer

CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats

CW153 – Common Wombat number 153

DEPC – Diethyl pyrocarbonate

dsDNA – double stranded DNA

EGK – Eastern Grey Kangaroo

eha – energy-conserving hydrogenase

ELISA – enzyme-linked immunosorbent assay

Env – Environment-associated

Fdh – formate hydrogenase

FDR – False discovery rate

Fla – flagellar/flagellin family protein

fwd – formylmethanofuran dehydrogenase

FXR – Farnesoid X receptor

GC – gas chromatograph

gDNA – Genomic DNA

GH – glycoside hydrolase

GIT – Gastrointestinal Tract

GT – glycosyltransferase

GTDB – Genome Tree Database

H_2 – hydrogen gas

H_2O – water

HA – Host-associated

HEK – human embryonic kidney

HIV – human immunodeficiency virus

HPLC – High-performance liquid chromatography

HQ – High-quality

Hsdh – hydroxysteroid dehydrogenase

IBD – Inflammatory bowel disease

IBS – irritable bowel syndrome

IBS-C – constipation-predominant IBS

IBS-D – diarrhoeal-predominant IBS

IgG – Immunoglobulin G

IMG – Integrated Microbial Genomes

IMO – Intestinal methanogen overgrowth

IPEC-J2 – intestinal porcine enterocytes isolated from the jejunum

iTOL – Interactive Tree of Life

JTT – Jones-Taylor-Thornton

Kb – kilo bases

Kbp – kilo basepair

KEGG – Kyoto Encyclopedia of Genes and Genomes

KO – KEGG Orthology

kPa – kilo pascal

LC-QToF – liquid chromatography quadrupole time of flight

MAG – Metagenome-assembled genome

MAMP – microbe-associated molecular pattern

Mbp – mega base pair

mcrA – methyl coenzyme M reductase subunit A

MetS – metabolic syndrome

MG – Mahogany Glider

MGS – metagenomic sequencing

mgs – methylamine-glutamate N-methyltransferase

mL – millilitre

mm – millimetre

mM – Millimolar

moDCs – monocyte-derived dendritic cells

MQ – Medium-quality

mRNA – messenger RNA

MtaB – methanol:cobalamin methyltransferase subunit B

mtr – tetrahydromethanopterin S-methyltransferase

N/S – not specified

NAFLD – non-alcohol fatty liver disease

NCBI – national centre for biotechnology information

nif – nitrogen fixation gene

nm – nanometres

NOD1/2 – Nucleotide-binding oligomerization domain-containing protein ½

OD_{600} – optical density at 600nm

ORF – open reading frame

OUT – Operational Taxonomic Unit

PBMC – Peripheral blood mononuclear cells

PBQC – Pooled biological quality control

PCA – Principal Component Analysis

PCR – polymerase chain reaction

Pva – penicillin v-acylase

qPCR – quantitative polymerase chain reaction

RBB+C – repeated bead beating plus column

RFLP – Restriction fragment length polymorphisms

RIN – RNA integrity number

RNA – ribonucleic acid

rpm – revolutions per minute

rRNA – ribosomal ribonucleic acid

RT-qPCR – Quantitative reverse transcription PCR

SBS – Short Bowel Syndrome

SDS – sodium dodecyl sulphate

SIBO – small intestinal bacterial overgrowth

SRA – sequence read archives

T2D – Type 2 Diabetes

TE – Tris-EDTA solution

TEM – transmission electron microscopy

Tg – trillion grams

TLR – toll-like receptor

TMA – trimethylamine

TMAO – trimethylamine *N*-oxide

TMAU – trimethylaminuria

TNFα – Tumour necrosis factor alpha

tRNA – Transfer RNA

UC – Ulcerative Colitis

V – volts

vol – volume

vWF – von Willebrand factor

Walc – *Methanosphaera* sp. WGK6 alcohol dehydrogenase

Wald – *Methanosphaera* sp. WGK6 aldehyde dehydrogenase

Chapter 1: Literature Review

In recent years, there has been an increased focus on the human microbiome and the role it plays in health and disease. These studies have largely focused the bacterial community of the human gastrointestinal tract (GIT) and the changes that result from variations in health and disease (Gilbert et al., 2018; Human Microbiome Project, 2012; Lurie-Weinberger & Gophna, 2015). However, improvements in sampling and microbial sequencing techniques are providing opportunities for the characterisation of microbes of "low abundance". The Archaea, dominated by the methanogenic Archaea (methanogens) in the human GIT, are a prime example of such microbes.

1.1 Human archaea

There is growing evidence that Archaea are both diverse and prevalent across multiple human body sites, as shown in Figure 1.1. The use of shotgun metagenomic sequencing with gut (stool) microbiota have revealed members of the *Euryarchaeota* phylogenetically affiliated with the *Archaeoglobales, Halobacteriales, Thermococcales* and *Thermoplasmatales,* along with representatives of the Crenarchaea including *Crenarchaeales, Sulfolobales, Desulfurococcales* and *Thermoproteales* (Gill et al., 2006; Nam et al., 2008; Rieu-Lesme et al., 2005). Interestingly, Hoffmann et al. (2013) identified *Nitrosphaera* lineages from stool samples of healthy subjects, but only in those samples that showed no detectable presence of *M. smithii*. Koskinen et al. (2017) reported that the bronchoalveolar lavage fluid (BAL) from healthy individuals was dominated by *Woesearchaeota* and other unclassified members of the DPANN superphylum, with *Thaumarchaeota* also detected. Archaea of the skin appear to be dominated by *Thaumarchaeota*, as observed from two separate studies on torso and torso/forearm/back samples, respectively (Koskinen et al., 2017; Moissl-Eichinger et al., 2017). This increased abundance of *Thaumarchaeota* observed from skin samples may be attributed to most of these species requiring oxygen for ammonia-oxidization (Hatzenpichler, 2012). These studies also observed the presence of *Woesearchaeota, Aenigmarchaeota* (DPANN), and *Halobacteriales*. A more in-depth analysis on the nasal microbiome of healthy men and women showed four archaeal phyla to be present: *Thaumarchaeota*, *Euryarchaeota*, *Pacearchaeota* and unclassified archaea, where *Thaumarchaeota* were represented by *Nitrososphaera* spp. and *Euryarchaeota* were represented by *Methanosphaera* spp., *Methanobrevibacter* spp. and *Halogranum* spp. (Koskinen et al., 2018). *Halobacteriales*-related sequences have been detected in faecal and intestinal biopsy samples across several studies (Gill et al., 2006; Nam et al., 2008; Rieu-Lesme et al., 2005). Oxley et al. (2010) showed the selective enrichment of *Halobacteriales* from human mucosal samples, providing evidence for their potential viability and persistence, which was further confirmed by the isolation of *Haloferax massiliensis* from the human GIT (Khelaifia & Raoult,

2016). Another representative, *Haloferax alexandrines,* was also recently isolated from human gut samples, giving further evidence for the presence of viable halophilic archaea within the gut (Khelaifia et al., 2017).

It is notable that the expansion of human archaeal diversity has primarily been obtained from DNA sequence data and is not readily "matched" with the recovery and isolation of cultured representatives of these "novel" lineages. Detection via DNA sequencing methods alone and/or via singular rather than multiple datasets are likely to be an inaccurate indicator of colonisation and persistence. For example, sequences of *Halobacteriales, Methanobrevibacter* and *Thermococci* have been detected

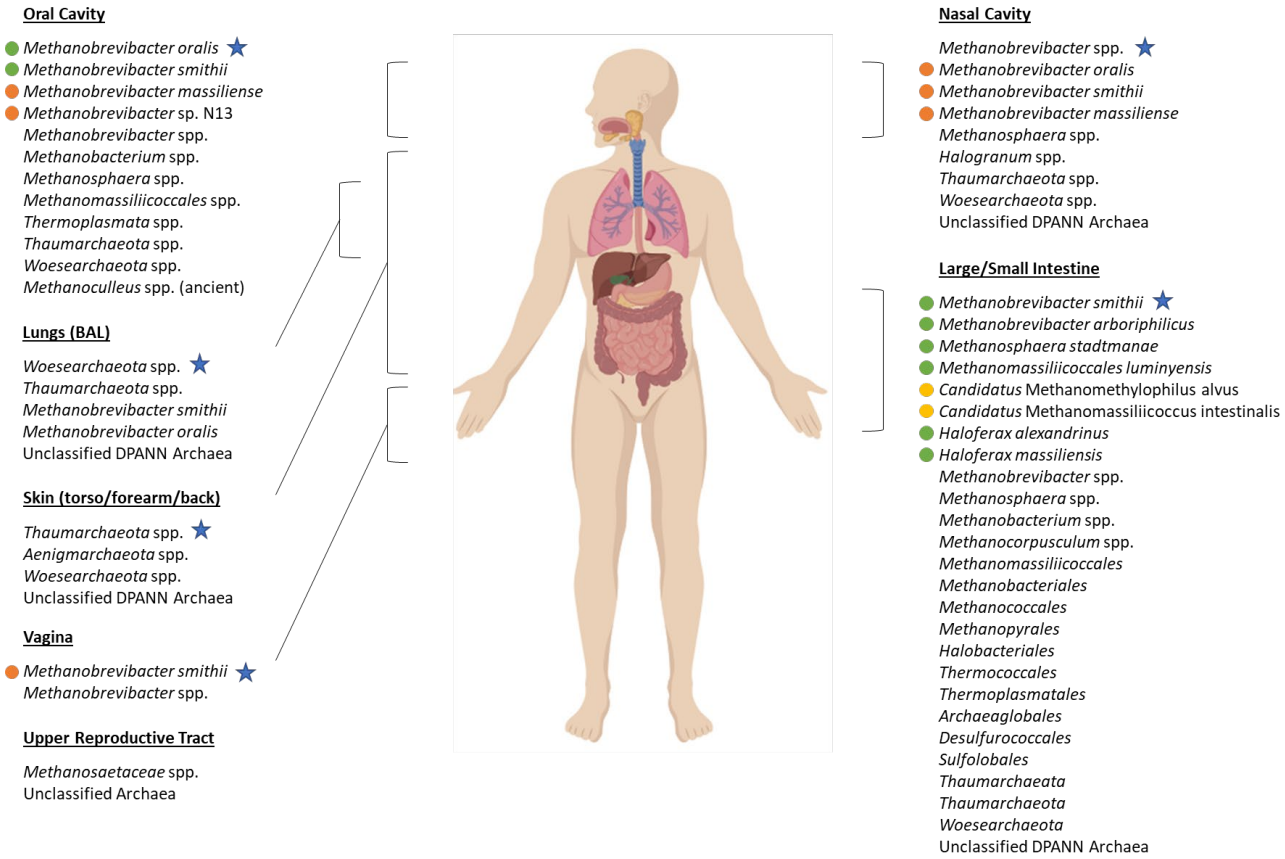


Figure 1.1. General human anatomical locations of isolated or detected archaeal species. Samples included those from the oral cavity, nasal cavity, lungs (BAL), skin (torso/forearm/back), large/small intestine, vagina and upper reproductive tract. ★ represents the most dominant archaea for a give sample region, ● represents axenic isolates with available genomic data, ● represents non-axenic enrichments with available genomic data, ● represents isolates with no available genomic data, and no symbol represent identification through sequencing data alone. The anatomical diagram created with BioRender. The figure was reworked and updated based on Bang and Schmitz (2015), and Nkamga et al. (2017), using additional information by Koskinen et al. (2017) and Hassani et al. (2020).

within salt, salt-preserved fermented seafood, and fish (Kobayashi et al., 2000; Nam et al., 2008). Additionally, archaeal species have been identified in bioaerosols, suggestion their detection from skin, nasal, and/or oral samples may be a reflection of inhalation and/or food ingestion, rather than active colonisation (Nehmé et al., 2009). In summary, while these findings are interesting, greater efforts need to be made to translate these newly identified archaea into a biological context, to determine whether they are indeed inherent members of the autochthonous human microbiota, or transiently detected because of environmental exposure and/or food-borne sources.

1.2 Methanogenic archaea

Unlike non-methanogenic archaea, methanogenic archaea are widely recognised as persistent members of the human gut microbiome. All currently recognised methanogen species belong to the phylum *Euryarchaeota*, historically classified into six distinct orders: *Methanobacteriales*, *Methanocellales*, *Methanococcales*, *Methanomicrobiales*, *Methanopyrales* and *Methanosarcinales*. In 2012, the most recent seventh order of methanogens was first proposed based on the isolation of the novel human isolate *Methanomassiliicoccus luminyensis* (Dridi, Fardeau, et al., 2012), later classified under the family *Methanomassiliicoccaceae* and order *Methanomassiliicoccales* (Borrel et al., 2014; Iino et al., 2013). Phylogenetic analysis of 16S ribosomal RNA (rRNA) genes separated the known methanogens into two unique classes (Bapteste et al., 2005); but further analysis using seven core cofactor biosynthesis and methanogenesis proteins provided further separation into Class I (*Methanobacteriales*, *Methanococcales*, *Methanomicrobiales*), Class II (*Methanomicrobiales*), and Class III (*Methanosarcinales, Methanocellales*) (Anderson et al., 2009). With the recently classified *Methanomassillicoccales* and candidate halophilic "*Methanonatronarchaeia"* (Sorokin et al., 2017), differentiation of methanogens by the current 'Class' system requires further revision.

1.3 The current paradigm - methanogenesis is an obligatory requirement for methanogen growth

Methanogenesis is the sole metabolic pathway for energy production in all known methanogenic archaea. This process involves the reduction of a carbon source to methane. The membrane-bound enzymes that catalyse this redox reaction are coupled to ion (proton) translocation across the cytoplasmic membrane (Welte & Deppenmeier, 2011), which produces an electrochemical gradient that drives the synthesis of ATP by A_1A_O-type ATP synthase (Deppenmeier & Muller, 2008; Pisa et al., 2007). Methanogens can be separated into three groups based on their substrate specificity: Hydrogenotrophic, Methylotrophic and Acetoclastic (Garcia et al., 2000), which are summarised in Figures 1.2–1.3. Hydrogenotrophic methanogenesis utilises hydrogen to reduce carbon dioxide to methane, with some species additionally able to utilise formate. This form of methanogenesis is

distributed across all three classes of methanogens, except for the *Methanomassiliicoccales* and candidate "*Methanonatronarchaeia"*. However, unlike class II and III methanogens, several class I methanogens are restricted exclusively to this pathway, such as *Methanobacterium bryantii* (Gilmore et al., 2017) and *Methanobrevibacter smithii* (Miller et al., 1982). Methylotrophic methanogenesis typically involves the hydrogen-dependent reduction of range of methylated compounds. *Methanosphaera stadtmanae* (Class I) performs the hydrogen-dependent reduction of methanol, where species of *Methanomassiliicoccus* and *Methanosarcina* (Class III) are able to additionally utilise methylated-amines (Borrel, Harris, et al., 2013; Ferry, 1999; Gorlas et al., 2012). Acetoclastic methanogenesis involves the production of methane from acetate and hydrogen, as performed by *Methanosarcina* spp. (Ferry, 1997)

The abovementioned substrates represent the classical examples for each of the three methanogenesis pathways. However, with advancements in genome sequencing and functional prediction, the metabolic versatility of methanogens is predicted to be greater than previously thought. For instance, *M. smithii* has long been thought to be restricted to growth with carbon dioxide (or formate) and hydrogen. However, Samuel et al. (2007) showed that co-culture of *M. smithii* and *B. thetaiotaomicron* in gnotobiotic mice resulted in the decreased concentration of caecal ethanol, as well as an increase in the expression of methanol:cobalamin methyltransferase B (*mtaB*) and an NADP-dependent alcohol dehydrogenase (*adh*). This suggested the potential of *M. smithii* to utilise ethanol in co-culture with B. *thetaiotaomicron* though biochemical validation is required to determine whether it is utilised as a substrate in methanogenesis (Samuel et al., 2007). Furthermore, *Methanococcoides* and *Methanosarcina* spp. use a variety of methylated thiols and amines, choline, glycine betaine and tetramethylammonium (Ticak et al., 2015; Watkins et al., 2012). The Morrison group isolated *Methanosphaera* sp. WGK6 from foregut digesta of a western grey kangaroo (*Macropus fuliginosus*) and showed this strain used ethanol in place of hydrogen gas for methanol reduction to methane. As a part of this work by Hoedt et al. (2016), *Methanosphaera* sp. WGK6 was shown to encode alcohol and aldehyde dehydrogenase genes for the predicted conversion of ethanol to acetaldehyde and subsequently acetate, with hydrogen generated utilise in methanol-dependent methanogenesis. Homologs of these genes were also found in multiple strains of *Methanobrevibacter*, suggesting this metabolic versatility may be widespread across methanogen lineages from different animals. Similarly, there are reports that *Methanocorpusculum* spp. use propanol, butanol, and pentanol as reducing agents for CO_2-dependent methanogenesis (Gilmore et al., 2017; Zellner et al., 1989).

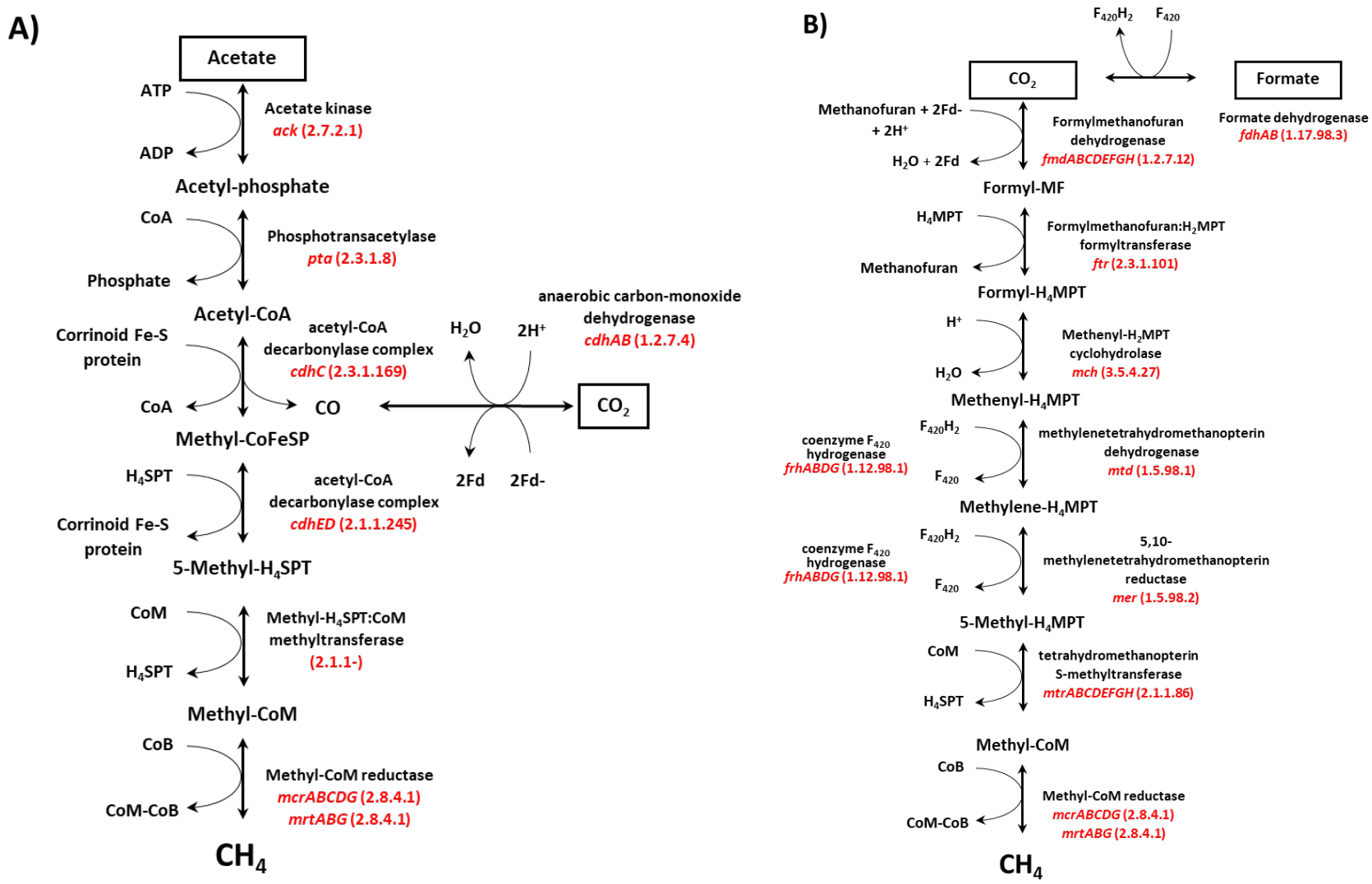


Figure 1.2. Pathways of hydrogenotrophic and acetoclastic methanogenesis. A) Acetoclastic pathway utilising acetate, **B)** hydrogenotrophic pathway utilising carbon dioxide (or formate). The black boxes represent the primary substrates utilised by each pathway of methanogenesis. This figure was adapted from Gilmore et al. (2017), with additions based on available methanogenesis KEGG pathways (Minoru Kanehisa et al., 2016).

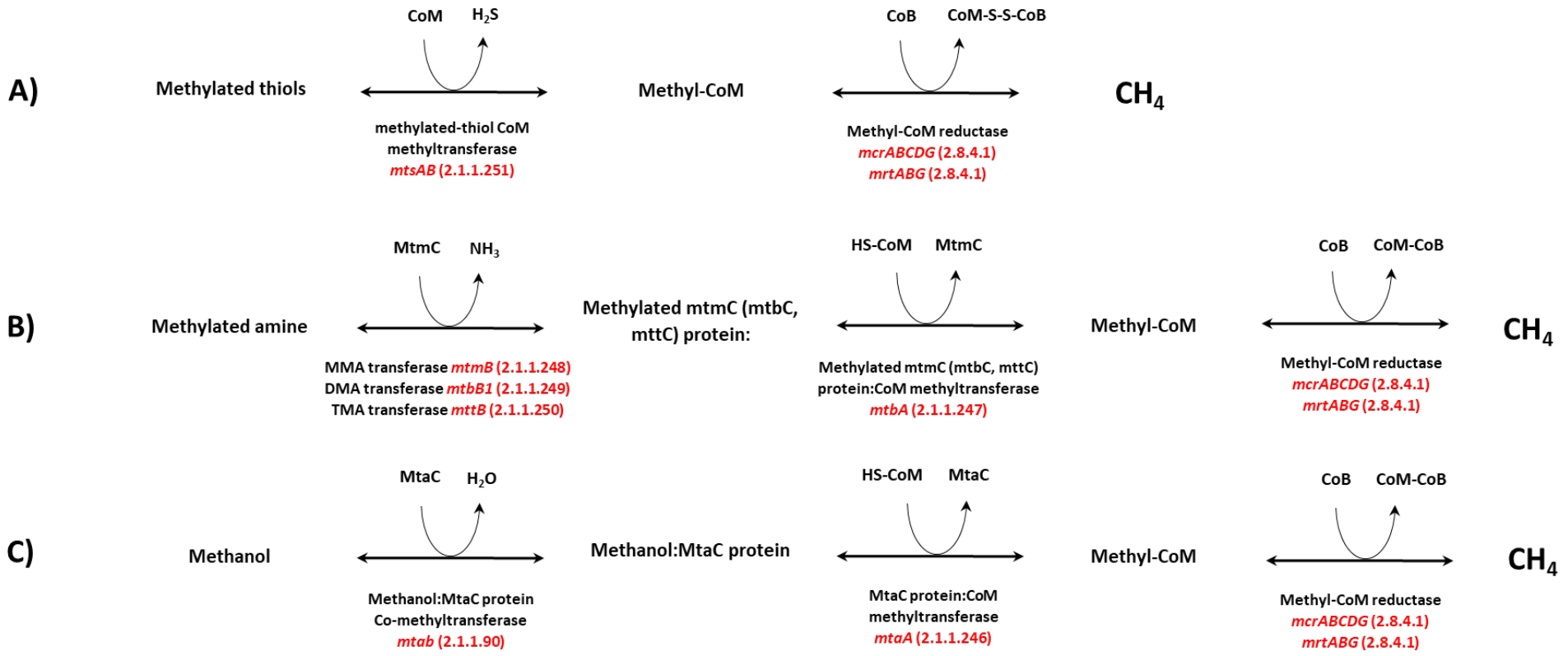


Figure 1.3. Pathways of common methylotrophic methanogenesis. A) Pathway utilising methylated thiols, **B)** Pathway utilising methylated amine, **C)** Pathway utilising methanol. The black boxes represent the primary substrates utilised by each pathway of methanogenesis. This figure was based on Gilmore et al. (2017), with additions based on available methanogenesis KEGG pathways (Minoru Kanehisa et al., 2016).

1.4 Human methanogenic archaea

The human methanogenic archaea are primarily represented by the *Methanobacteriales* and *Methanomassiliicoccales* (Chaudhary et al., 2015). The *Methanobacteriales* populations are typically dominated by *M. smithii,* with their prevalence approaching 95% (Dridi, 2012; Dridi et al., 2009; Dridi, Raoult, et al., 2011) and account for as much as 10% of the total microbiome in some studies (Eckburg et al., 2005). *Methanosphaera stadtmanae* is less common and abundant, found in ~30% of individuals (Dridi, Henry, et al., 2012). The *Methanomassiliicoccales,* represented by *Methanomassiliicoccus luminyensis,* are the least common, although there are some reports of high prevalence (4%-50% of individuals tested), and their relative abundance increases with age (Dridi, Henry, et al., 2012; Vanderhaeghen et al., 2015). The use of cultivation methods to enrich for specific methanogens has also resulted in the identification of *Candidatus* Methanomethylophilus alvus (Borrel et al., 2012) and *Candidatus* Methanomassiliicoccus intestinalis (Borrel, Harris, et al., 2013). Methanogens have also been identified among the communities of other body sites, with *Methanobrevibacter* (*oralis*) and *Methanomassiliicoccales* spp. detected in the oral cavity (Horz et al., 2012; Li et al., 2009). Although phylogenetically similar, *M. oralis* and *M. smithii* show adaptations to their respective biological niches, with *M. oralis* isolates lacking the capacity to utilise formate (Ferrari et al., 1994) and *M. smithii* encoding for bile salt hydrolase (*bsh*) genes (Gaci et al., 2014). Other members of the *Methanobrevibacter* genus found in humans include *M. arboriphilicus* (Khelaifia et al., 2014) and *M. massilience* (Huynh et al., 2017), but there is currently little information about their respective prevalence or abundance. Interestingly, cultivation of oral methanogens from three individuals with severe periodontitis identified a novel *Methanobrevibacter* species designated N13, along with multiple representatives of *M. smithii* and *M. oralis* (Huynh et al., 2015).

Like the non-methanogenic archaea, there are a number metagenomic sequencing studies that suggest a greater diversity of low abundance methanogens. Representatives of the *Methanobacteriales, Methanomicrobiales, Methanococcales, Methanopyrales,* and *Methanosarcinales* have been identified via shotgun metagenomics sequencing (Bang & Schmitz, 2015; Scanlan et al., 2008). Additionally, studies have also uniquely identified methanogen species from mucosal samples. Along with the abovementioned *M. arboriphilus,* methyl coenzyme M reductase A (*mcrA*) clone sequences closely related to *Methanoculleus chikugoensis* were found, along with oral representatives of *M. congolense* and *M. mazei* (Nava et al., 2012; Nguyen-Hieu et al., 2013). A separate study on longitudinal GIT biopsies retrieved *Methanobrevibacter* sequences related to *M. filiformis* and *M. woesei*, along with the first identification of *Methanobacterium* sequences specifically within the ileum (Koskinen et al., 2017). This study additionally showed *Methanobacteriaceae* present within

nasal samples. Methanogenic archaea identified as *M. smithii* have been detected in vaginal samples of individuals suffering from bacterial vaginosis (Belay et al., 1990; G. Grine et al., 2019). Little was known about archaea in the reproductive tract of healthy individuals until a recent study by Li et al. (2018) showed *Methanosaetaceae* within cervical mucus and peritoneal fluid samples. In summary, although typically dominated by *Methanobrevibacter* spp., a wide variety of methanogens colonise the human body, comprising a low abundance archaeome that is not currently recognised based on faecal microbiome sequencing or cultured isolation. Further work is needed to explore this low abundance methanogen community, and the role it collectively plays in the larger microbiome and GIT health.

1.5 Methanogenic archaea in health and disease

There has been a gradual but sustained increasing interest in archaea, specifically methanogens, and their relationship with human health and disease. Table 1.1 provides a summary of the associations between the relative and/or absolute abundance of methanogenic archaea with different non-communicable diseases. Despite these associations, there is scant biology to dissect causality from these methanogens and their host interactions. In the following sections, I provide a brief overview of these interrelationships and highlight some of the knowledge gaps for improving our understanding of the roles of methanogenic archaea in health and disease.

Table 1.1. Associations between methanogenic archaea and different diseases or disorders. Upward arrows (↑) and downward arrows (↓) represent changes of methanogen abundance and breath-methane excretion, for the respective studies. IBD, Inflammatory bowel disease; CD, Crohn's Disease; UC, Ulcerative Colitis; IBS, Irritable bowel syndrome; IBS-C, constipation-predominant IBS; IBS-D, diarrhoeal-predominant IBS; SBS, Short Bowel Syndrome; MetS, metabolic syndrome; SIBO, small intestinal bacterial overgrowth. * denotes detection within respective sample types.

Pathology	Association	Method	Citation
IBD - CD	↑ *Methanosphaera* ↑ Msp-specific IgG	RT-qPCR (*MtaB1*) Indirect ELISA	(Blais Lecours et al., 2014)
	↓ *Methanobrevibacter*	Metagenomic shotgun sequencing	(Lo Sasso et al., 2020)
	↓ *Methanobrevibacter*	16S rRNA Sequencing	(Pascal et al., 2017)
IBD - CD/UC	↓ Methane production	Breath-methane test	(Peled et al., 1987)
	↓ Methane production	Breath-methane test	(McKay et al., 1985)
	↓ Methanogens	PCR (*mcrA*)	(Scanlan et al., 2008)
	↑ *Methanobrevibacter*	16S rRNA RT-qPCR	(Verma et al., 2010)
	↓ *Methanobrevibacter* ↑ *Methanobrevibacter* in remission	16S rRNA RT-qPCR	(Ghavami et al., 2018)
	↓ Methane Metabolism	16S rRNA PICRUSt predictions	(Nishino et al., 2018)

	↓ *Methanobrevibacter*	RT-qPCR	(Heidarian et al., 2019)
	↓ Methane production	Breath-methane test	(Gu et al., 2020)
IBS	No sig. association	Breath-methane test	(Bratten et al., 2008)
	No sig. association	PCR (*mcrA*)	(Scanlan et al., 2008)
IBS - Constipation	↑ Methane production ↑ Methane with severity	Breath-methane test	(Pimentel et al., 2003)
	↑ Methane production ↑ *M. smithii*	Breath-methane test 16S rRNA RT-qPCR	(Ghoshal et al., 2016)
	↑ Transit time	Breath-methane test	(Attaluri et al., 2010)
	↑ Transit time	Breath-methane test	(Lee et al., 2013)
	↑ Transit time	Breath-methane test	(Pimentel et al., 2006)
	↑ Transit time	Breath-methane test	(Majewski & McCallum, 2007)
	↑ Transit time	Breath-methane test	(Hwang et al., 2010)
	↑ Transit time	Breath-methane test	(Ghoshal et al., 2011)
	↑ Methane with severity	Breath-methane test	(Chatterjee et al., 2007)
	↑ *Methanobrevibacter*	16S rRNA Sequencing	(Pozuelo et al., 2015)
	↑ *M. smithii* ↑ Methane production	16S rRNA Sequencing Breath-methane test	(Kim et al., 2012)
IBS - Diarrhoea	↓ Methane production	Breath-methane test	(Pimentel et al., 2003)
	↓ *Methanobacteriales*	16S rRNA RT-qPCR	(Tap et al., 2017)
Diverticulosis	↑ Methane production ↑ Methanogen abundance	Breath-methane test Culture-based	(Weaver et al., 1986)
Colorectal Cancer	↑ Methane with severity	Breath-methane test	(Pique et al., 1984)
	↑ Methane production	Breath-methane test	(Haines et al., 1977)
	↓ Methanogen abundance	Breath-methane test	(Segal et al., 1988)
	↓ Methanogen abundance	Metagenomic shotgun sequencing	(Coker et al., 2020)
	↑ *Methanobacteriales* ↑ *Methanobrevibacter*	16S rRNA qPCR	(Mira-Pascual et al., 2015)
	No sig. association	Breath-methane test	(O'Keefe et al., 2007)
	No sig. association	Breath-methane test	(Segal et al., 1988)
	No sig. association	Breath-methane test	(Karlin et al., 1982)
	No sig. association	Breath-methane test	(Kashtan et al., 1989)
	No sig. association	Breath-methane test	(Hoff et al., 1986)
	No sig. association	PCR (*mcrA*)	(Scanlan et al., 2008)
Obesity	↑ Methanogen abundance	16S rRNA Pyrosequencing	(Zhang et al., 2009)
	↓ *M. smithii*	16S rRNA RT-qPCR	(Million et al., 2012)
	↓ *M. smithii*	16S rRNA RT-qPCR	(Million et al., 2013)
	↓ *M. smithii* ↑ Unclassified *Methanobrevibacter*	Metagenomic shotgun sequencing	(Maya-Lucas et al., 2018)
	↑ Methane production	Breath-methane test	(Mathur et al., 2013)
	No sig. association	16S rRNA RT-qPCR	(Schwiertz et al., 2010)
	No sig. association	Breath-methane test 16S rRNA RT-qPCR	(Fernandes et al., 2013)

Anorexia	↑ *M. smithii*	16S rRNA RT-qPCR	(Armougom et al., 2009)
	↑ *M. smithii*	RT-qPCR	(Million et al., 2013)
	↑ *M. smithii*	16S rRNA RT-qPCR	(Borgo et al., 2017)
	↑ *Methanobrevibacter*	16S rRNA Sequencing	(Mack et al., 2016)
	↑ *M. smithii*	16S rRNA Sequencing	(Prochazkova et al., 2021)
Malnutrition	↓ *M. smithii*	Archaeal-specific qPCR	(Million et al., 2016)
	↓ *M. smithii*	16S rRNA RT-qPCR	(Camara et al., 2021)
	↑ *M. smithii*	16S rRNA Sequencing	(Kamil et al., 2021)
Encopresis	↑ Methane production	Breath-methane test	(Fiedorek et al., 1990)
SBS	↓ *M. smithii*	16S rRNA RT-qPCR	(Boccia et al., 2017)
Mets	↑ *Methanobrevibacter*	16S rRNA Sequencing	(Lim et al., 2017)
Multiple Sclerosis	↑ *Methanobrevibacter* ↑ Breath methane	16S rRNA Sequencing Breath-methane test	(Jangi et al., 2016)
	↑ *Methanobrevibacter*	16S rRNA Sequencing	(Tremlett, Fadrosh, Faruqi, Zhu, et al., 2016)
	↑ *Methanobacteriaceae*	16S rRNA Sequencing	(Jhangi et al., 2014)
Parkinson's Disease	↑ Methanogen abundance	16S rRNA Sequencing	(Qian et al., 2018)
	↑ Methanogen abundance	Metagenomic shotgun sequencing	(Bedarf et al., 2017)
SIBO	↑ Breath methane	Breath-methane test	(Suri et al., 2018)
	↑ Breath methane	Breath-methane test	(Pimentel et al., 2020)
Vaginosis	↑ *M. smithii*	Antigenic fingerprinting	(Belay et al., 1990)
	↑ *M. smithii*	16S rRNA Sequencing	(G. Grine et al., 2019)
Urinary Tract Infection	↑ *M. smithii*	16S rRNA/McrA PCR RT-qPCR/Culture	(Ghiles Grine et al., 2019)
Anerobic Abscesses	*M. smithii* *	16S rRNA/McrA Sequencing	(Nkamga et al., 2016)
	↑ *M. oralis*	16S rRNA RT-qPCR/ Metagenomic sequencing	(Drancourt et al., 2017)
	M. oralis *	16S rRNA/McrA Sequencing	(Nkamga et al., 2018)
Periodontal Disease	↑ *M. oralis* ↑ *M.* sp. strain N13	16S rRNA/McrA Sequencing/Culture	(Huynh et al., 2015)
	M. massiliense *	Culture	(Huynh et al., 2017)
	↑ *M. oralis*	Archaeal PCR	(Li et al., 2009)
Parasitic Infection	↑ *Methanobrevibacter*	16S rRNA Sequencing	(H. Chen et al., 2021)

Cirrhosis	↓ *Methanobrevibacter*	16S rRNA Sequencing	(Ponziani et al., 2021)

1.6 Methanogenic archaea and gut nutritional ecology

In terms of maintaining health, methanogens play an important role as the terminal step in bacterial fermentation, where the carbon dioxide and hydrogen gas are utilised by methanogens for methanogenesis. This process is known as interspecies hydrogen (and carbon dioxide) transfer and serves to limit the build-up of hydrogen gas, which can inhibit bacterial fermentation and growth (Nakamura et al., 2010). The removal of these end products conserves the thermodynamic equilibrium of fermentation, maintaining 'microbial homeostasis' within the human GIT (Sieber et al., 2012; Stams & Plugge, 2009). For *Methanosphaera*, a source of methanol is necessary for growth, which can come in the form of free methanol (spirits, beer, wine), methyl esters of fatty acids (aspartame) and pectin (fruit and vegetable)(Toxicity, 2011). There is currently no evidence to suggest *Methanosphaera* can utilise pectin directly, so the degradation of pectin by pectinase-containing bacteria, such as *Bacteroides* spp., is necessary for methanol availability (Dongowski et al., 2000; Jensen & Canale-Parola, 1986). For *Methanomassiliicoccales* spp., methylated-amines produced from dietary carnitine, choline, trimethylamine *N*-oxide (TMAO) and phosphatidylcholine from meat, eggs, nuts, and fish can be utilised. These compounds are broken down by resident microbial communities, such as the conversion of TMAO to trimethylamine (TMA) by *Enterobacteriaceae* spp., to produce free methylated amines (Hoyles et al., 2018; Rebouche & Chenard, 1991; Spencer et al., 2011; Tang et al., 2013; Zeisel et al., 1983). Comparatively, there is little information of the role non-methanogenic archaea play in nutritional ecology of the gut. Given that halophilic archaeal sequences are frequently identified in high salt food products, it is reasonable to assume a portion of halophilic archaeal load may be directly associated with dietary intake (Kobayashi et al., 2000). However, some species of *Halobacteriaceae* are able to survive in salt concentrations similar to that of average salinity levels of healthy individuals (~140 mM sodium) (Fukushima et al., 2007). In addition, small pockets of concentrated luminal ions have also been identified within the colon, potentially acting as favourable micro-niches for these organisms (Naftalin & Pedley, 1995; Spring, 1998). Halophilic archaea have also demonstrated the ability of survive under anaerobic conditions, utilizing electron acceptors such as fumarate for the fermentation of compounds such as arginine (Oxley et al., 2010). Oxley et al. (2010) also noted the increase in luminal osmolality and organic solute concentration of IBD patients as a potential factor for the increase in halophilic archaea. Despite these linkages, the ecological and metabolic niche that halophilic archaea occupy within the GIT is currently inferential, but with the isolation of human *Haloferax* spp. (Khelaifia et al., 2017; Khelaifia & Raoult, 2016), there is now an opportunity to better

define the nutritional ecology of these organisms within the human gut. In fact, a recent study on the bacterial and archaeal composition of colorectal cancer patients showed an increased presence of the halophilic *Natrinema* sp. J7-2 and concurrent reduction in methanogens compared to control subjects (Coker et al., 2020). Additionally, the characterisation of the archaeal community of South Korean individuals showed 42.47% archaeal positivity, with 95.54% of archaeal-positive faecal samples containing haloarchaea-associated sequences (Kim et al., 2020). Although the average relative abundance of haloarchaea species was 9.63%, some individuals within the cohort displayed a haloarchaea-dominant archaeal community with up to 99.33% relative abundance (Kim et al., 2020).

1.7 Methanogenic archaea and gastrointestinal motility

Although there are currently no conclusive findings on the role of methanogenic archaea in human disease, there have been numerous associations made to intestinal-associated pathologies. Breath methane has historically been used to test for the presence of methanogens prior to the development of next-generation sequencing techniques. As summarised by de Lacy Costello et al. (2013), this technique involves the ingestion of a sugar, typically lactose, glucose or fructose, and analysis of alveolar methane over the subsequent 1-2 hour period. An increase in the breath methane of constipation-predominant IBS (IBS-C) patients has been associated with increase severity and increase intestinal transit time (Chatterjee et al., 2007; Pimentel et al., 2003). Additionally, an increase in *Methanobrevibacter*, specifically *M. smithii*, has been associated with IBS-C by 16S rRNA sequencing (Ghoshal et al., 2016; Pozuelo et al., 2015). Conversely, individuals with diarrhoeal-predominant IBS show a reduction in both methane production and *Methanobacteriales* abundance (Pimentel et al., 2003; Tap et al., 2017). IBS broadly appears to have no significant association with breath methane or methanogen abundance, though failure to recognise and separate IBS-C/D patients may provide an explanation for these findings (Bratten et al., 2008; Scanlan et al., 2008).

Small intestine bacterial overgrowth (SIBO) is a symptom associated with IBD/IBS patients, in which there is a significant increase in small intestinal bacteria (Colombel et al., 2018). SIBO is relatively common in patients with UC, with ~30% presenting with the condition, compared to a lower prevalence observed in patients with CD (Lee et al., 2015; Sandborn, 2009). A recent study by Suri et al. (2018) showed delayed motility in SIBO to correlate with an increase in breath-methane levels. In a separate study on patients IBD and SIBO, individuals categorised under IBS-C were more likely to be methane producers compared to IBS-D (58% compared to 28% (Majewski & McCallum, 2007). Comparatively, individuals with IBS-D were more likely to be hydrogen producers. Given the implication of methanogens and methane in SIBO, recent recommendations by the American College of Gastroenterology include the terminology of intestinal methanogen overgrowth (IMO) to better

represent the overgrowth of methanogens in the small intestine and colon (Pimentel et al., 2020). In fact, methane itself has been linked to a reduction in intestinal transit frequency. Jahng et al. (2012) used sections of guinea pig ileum submerged in a peristaltic bath to show an infusion of methane caused decreased peristaltic velocity and increased contraction amplitude, compared to increased peristalsis with hydrogen gas. Additionally, hydrogen was also shown to decrease transit time by 47% in the proximal colon (Jahng et al., 2012). This suggests a possible positive feedback loop between methanogen growth, methane production and increased retention times, caused by a neuromuscular transmitter-like effect of methane (Furnari et al., 2012; Triantafyllou et al., 2014).

Contradictory associations are observed in obese individuals, with an overall increase in the methanogen population but a shift away from *M. smithii* towards unclassified *Methanobrevibacter*, though other studies show no significant association (Fernandes et al., 2013; Mathur et al., 2013; Maya-Lucas et al., 2018; Million et al., 2013; Million et al., 2012; Schwiertz et al., 2010; Zhang et al., 2009). Individuals with severe malnutrition show reduced *M. smithii* abundance, which may be explained by a lack of intestinal nutrients and thus bacterial fermentation (Million et al., 2016). Indeed, this was recently affirmed in patients with severe acute malnutrition, which showed *M. smithii* in only 4.2% of cases compared to 40.9% in control subjects (Camara et al., 2021). In contrast, individuals with anorexia show a significantly increase in *M. smithii* in multiple studies, as do individuals with metabolic syndrome, suggesting altered microbial communities could affect methanogen populations (Armougom et al., 2009; Borgo et al., 2017; Lim et al., 2017; Mack et al., 2016; Prochazkova et al., 2021).

1.8 Methanogenic archaea and infection?

While methanogens are historically characterised as commensal members of the gut microbiome, recent studies have provided evidence implicating methanogenic archaea in polymicrobial infections. For instance, 16S rRNA gene profiling studies have identified *Methanobrevibacter, Methanobacterium, Methanosarcina, Methanosphaera,* and *Thermoplasmatales* present in subgingival plaque (Belay et al., 1988; Horz et al., 2012; Kulik et al., 2001; Li et al., 2009; Robichaux et al., 2003). However, *M. oralis* is the only species to be significantly associated with periodontal disease, as summarised by Nguyen-Hieu et al. (2013). One of the most common treatment options for periodontitis is metronidazole and is one of the few widely used antibiotics with efficacy against methanogens such as *M. oralis* (Dridi, Fardeau, et al., 2011). Thus, the metronidazole-associated suppression of *M. oralis* may play a significant role in effective treatment (Nguyen-Hieu et al., 2013). Conversely, a separate study showed *M. oralis* to have no significant increase in prevalence for peri-implantitis, suggesting a potential for disease-specific associations in the oral microbiome (Belkacemi

et al., 2018). Specific amplification of archaeal 16S rRNA and *mcrA* showed *M. smithii* sequences in chronic paravertebral abscess of a 41-year-old man (Nkamga et al., 2016). The group was also able to isolate *M. oralis* from a nasal sample of a patient suffering from chronic sinusitis. There is also growing evidence that methanogens may contribute to disease progression of brain abscesses. One recent study by Drancourt et al. (2017) showed a higher prevalence of archaeal species by PCR in brain abscess specimens compared with healthy controls. Additionally, metagenomics analysis identified *M. oralis* within multiple abscess samples, as well as several bacterial species, including *S. intermedius*. Mice infected cerebrally with *M. oralis, S. intermedius* or both showed significantly increased mortality in all test cases compared to controls. Additionally, co-infection with *M. oralis* and *S. intermedius* showed an increased mortality rate compared to separate infections, suggesting a syntrophic relationship between the microbes. *M. oralis* was further observed in a community-acquired brain abscess of a 30-year-old woman along with *A. actinomycetemcomitans*, again suggesting a potential role for *M. oralis* in infections associated with anaerobic bacteria (Nkamga et al., 2018).

1.9 Interactions between methanogenic archaea and the immune system

Using a murine model of archaeal airway exposure, Blais Lecours et al. (2011) showed that nasal administration of both *M. smithii* and *M. stadtmanae* biomass induced alveolar accumulation of granulocytes and macrophages, as well as thickening of the alveolar septa. While the effects from *M. smithii* challenge were relatively mild, there was a much stronger response towards *M. stadtmanae*, and in a separate study *M. stadtmanae*-induced pneumonitis in mice also caused a significant induction of B-cell-rich tertiary lymphoid tissues (Huppe et al., 2018). When the recruitment of B-cells was prevented by an agonist of sphingosine-1-phosphate receptor 1, a key regulator of lymphoid cells, *M. stadtmanae*-specific lung antibody titres were reduced along with airway leakage and neutrophilic inflammation (Huppe et al., 2018). In a murine model of airway inflammation, crude *Methanosphaera* and *Methanobrevibacter* extracts induced a TH_{17}-dependent type IV hypersensitivity response (Bernatchez et al., 2017). Additionally, the *Methanosphaera*-specific immune response also presented with high titres of antigen-specific IgG_1 and IgG_{2a}, again showing the increased immunogenicity of *Methanosphaera* (Bernatchez et al., 2017). Collectively, these results suggest that human archaea, specifically *M. stadtmanae*, stimulate both arms of the immune system and induce a significant proinflammatory immune response. However, further work is needed to understand the archaea-induced inflammatory response in the progression and maintenance of gastrointestinal diseases such as IBD.

Bang et al. (2014) showed that *M. smithii* and *M. stadtmanae* were not recognised by Caco-2/BBe human epithelial cells, in terms of cytokine and antimicrobial peptide production, as was previously shown for intestinal commensal bacteria (Sansonetti, 2004). However, both *Methanosphaera* and *Methanobrevibacter* displayed a decreased growth rate and yield when exposed to a derivative of human cathelicidin, as well as a synthetic anti-lipoprotein peptide (Lpep) and porcine lysin NK-2, when supplemented in axenic culture (Bang et al., 2012). Similarly, *M. luminyensis* showed a high sensitivity to human cathelicidin, though it was significantly more resistant to Lpep and porcine lysin NK-2 (Bang et al., 2017). This mechanism was further explored by identifying specific Toll-like receptors (TLRs) for the recognition of methanogen-specific microbe-associated molecular patterns (MAMPs). Using human embryonic kidney (HEK) cells transfected with specific intracellular TLRs (3, 7, 8, and 9), Bang et al. (2014) showed no activation by RNA or DNA from heat-inactivated archaeal cell preparations. Similarly, no recognition was observed in TLR5 cells, which play an essential role in the recognition of flagella (Bang et al., 2014). This was not unexpected as there is no genetic evidence of flagellin-like genes within *M. stadtmanae* and only two predicted flagellin-like genes within the genome of *M. smithii* PS (Fricke et al., 2006; Samuel et al., 2007). TLR2, NOD1 and NOD2 were also tested for their role in the recognition of bacterial cell membrane components, such as lipid (TLR2) and murein (NOD1/NOD2) (Girardin, Boneca, Carneiro, et al., 2003; Girardin, Boneca, Viala, et al., 2003; Kataoka et al., 2006). Neither TLR2 nor NOD1/NOD2 cells displayed recognition of *M. smithii* and *M. stadtmanae*, suggesting the archaeal cell wall components are immunologically distinct from those of pathogenic bacterial species (Bang et al., 2014). However, contrary to these results, the stimulation of TLR knockout human monocyte BLaER1 cell lines showed not only *M. stadtmanae* itself but also preparations of *M. stadtmanae* RNA to elicit a TLR7- and TL8-specific immune recognition, with the latter showing a greater response (Vierbuchen et al., 2017). Additionally, the TLR8-specific response was able to induce the activation of the NLRP3 inflammasome (Vierbuchen et al., 2017). Thus, further work is warranted to better characterise the potential MAMPs of archaeal species and their associated TLR activation pathway. Despite this variation in response, both strains induced maturation of monocyte-derived dendritic cells (moDCs) through to up-regulation of CD197 and CD86 (Bang et al., 2014). Additionally, confocal and transmission electron microscopy (TEM) was used to show phagocytosis of the methanogens was required for activation of the moDCs (Bang et al., 2014). In a subsequent study, *M. luminyensis* showed a weak response in both moDCs and peripheral blood mononuclear cells (PBMCs), suggesting a lower immunogenic potential towards human immune cells (Bang et al., 2017).

1.10 Methanogenic archaea, IBD, and colorectal cancer

With breath methane testing, individuals with IBD show reduced methane expulsion (Gu et al., 2020; McKay et al., 1985; Peled et al., 1987). Scanlan et al. (2008) further validated these results by PCR of the methanogenesis marker gene *mcrA*. This successfully showed a reduction in the abundance of methanogens in individuals with IBD, with UC patients showing a 24% reduction and patients with CD showing a 30% reduction (Scanlan et al., 2008). Subsequently, patients with CD were showed a specific reduction of *Methanobrevibacter*, and a shift towards *Methanosphaera,* which may be responsible for the reduction in breath methane (Blais Lecours et al., 2014; Ghavami et al., 2018). Similarly, this was recently replicated in a population of Kazan IBD patients, which showed a significant reduction of Euryarchaeota, attributed to *Methanobrevibacter*, in patients with CD compared to those with UC (Lo Sasso et al., 2020). Methane metabolism has also been shown as reduced in patients with IBD compared to control subjects (Nishino et al., 2018). Blais Lecours et al. (2014) specifically showed an increased prevalence of *M. stadtmanae* in patients with IBD compared to control subjects. Additionally, it was also shown that IBD patients produced a significant *M. stadtmanae*-specific IgG immune response compared to non-IBD healthy individuals and PBMCs produced a higher proinflammatory cytokine (TNFα) response when exposed to *M. stadtmanae* compared to *M. smithii*. Stimulation of moDCs also showed *M. stadtmanae* to elicit a significant proinflammatory cytokine response compared to *M. smithii* (Bang et al., 2014).

However, a study on patients with UC and CD from an Indian population showed a converse shift in methanogens, with an increase observed in *Methanobrevibacter* for both patient groups compared to controls (Verma et al., 2010). Individuals with short bowel syndrome (SBS) due to surgical intervention show a decrease in the abundance of *M. smithii*, possibly due to the physical restriction of extended retention times (Boccia et al., 2017). Conversely, individuals with diverticulosis showed an increase in *Methanobrevibacter* compared to standard IBD patients, potentially due to the diverticula creating micro-niches for the methanogens within the colon (Weaver et al., 1986). There is scant information available on the role of non-methanogenic archaea gut disease. However, one study by Oxley et al. (2010) identified multiple unique *Halobacteriaceae* phylotypes within biopsy and faecal samples of patients with IBD, though there is currently no information to suggest what role, if any, these species play IBD.

Multiple studies show an increase of methanogens and methane production associated with colorectal cancer (CRC) and the stage of disease (Haines et al., 1977; Mira-Pascual et al., 2015; Pique et al., 1984; Segal et al., 1988). However, many studies also show no significant associations between the two, suggesting that the association between methanogens and CRC may involve complex factors that are currently not well understood (Hoff et al., 1986; Karlin et al., 1982; Kashtan et al., 1989;

O'Keefe et al., 2007; Scanlan et al., 2008). Although, a recent study on CRC patients showed an enrichment in haloarchaea and a concurrent reduction in methanogens compared to control subjects (Coker et al., 2020).

1.11 Gut archaea – systemic and metabolic disease

Euryarchaeota has also been implicated in autoimmune diseases with potential links to the microbiome, such as an increase associated with shorter relapse time for paediatric multiple sclerosis patients (Castillo-Alvarez et al., 2018; Tremlett, Fadrosh, Faruqi, Hart, et al., 2016) or a correlation to increased disease activity score in patients with rheumatoid arthritis (Picchianti-Diamanti et al., 2018). Adults with asthma were also found to have a reduction in *M. smithii* compared to control subjects (Wang et al., 2018). Individuals with metabolic syndrome (MetS) were observed to have an increase in *Methanobrevibacter* compared to control subjects (Lim et al., 2017).

Despite the correlation of methanogens to various diseases, little information is available on whether methanogens are playing an active role or are simply responding to ecological changes. Despite the hypothesised use of *Methanomassiliicoccales* as probiotics for the reduction for uraemic toxins, there are no significant associations observed between these species and atherosclerotic cardiovascular disease (Jie et al., 2017). Similarly, patients with chronic kidney disease (CKD) are also associated with increased levels of uraemic toxins, such as TMAO (Lau et al., 2018). However, analysis of CDK individuals by our group showed no increase in methanogens, including *Methanomassiliicoccales,* with TMAO concentrations (unpublished data). As such, the diseased environment may be unfavourable for these methanogens.

One important host-derived factor which is altered in disease states is bile salt. Bile salts are molecules with detergent-like properties secreted into the small intestine to aid in the digestion of dietary fats (Barrasa et al., 2013; Maldonado-Valderrama et al., 2011). Typically, most of the bile will be reabsorbed in the small intestine through enterohepatic circulation, however impaired intestinal function can result in an increase concentration of colonic bile (Barrasa et al., 2013; Boyer, 2013). As summarised by Joyce and Gahan (2017), intestinal diseases including IBD, SBS, IBS-C/D and *Clostridium difficile* infection (CDI) are associated with various changes in the bile salt pool. In the case of IBD, there is a significant reduction in bile salt deconjugation, desulphation and transformation reactions, along with a reduction in *bsh* activity (Duboc et al., 2013). This reduced activity is attributed to the reduction of bile salt metabolising bacteria, such as *Firmicutes*, resulting in a shift towards primary and sulphated bile salts (Duboc et al., 2013). Genomic analysis shows most GIT methanogen isolates to contain at least a single copy *bsh*, except for the *Methanomassiliicoccales* (Gaci et al., 2014). With regard to the most dominant methanogens, the *M. smithii* and *M. stadtmanae*

type strains both contain *bsh* (Fricke et al., 2006; Samuel et al., 2007). Indeed, transformation of the *bsh* of *M. smithii* (PS) into *E. coli* showed both tauro- and glyco-conjugated bile salt activity (Jones et al., 2008). Additionally, bioinformatics analyses have also shown both species to encode for predicted hydroxysteroid dehydrogenases (*hsdh*)(Doden et al., 2018; Kisiela et al., 2012). Thus, these methanogens have a genetic capacity to adapt to changing bile salt conditions such as those seen in IBD and warrants further investigation into the species-specific effects of bile salt. Additionally, this could also explain the relatively low abundance of *Methanomassiliicoccales* spp. within the human gut.

1.12 Gut archaea – probiotic potential?

With their specific substrate utilisation, there is an interest in using certain archaeal species as probiotic supplements, termed 'archaebiotics', in certain disease states (Brugere et al., 2014). The most plausible example is the potential for utilising *Methanomassiliicoccus* spp. as a probiotic therapy for trimethylaminuria (TMAU). This disease predominantly involves a genetic susceptibility arising from mutation(s) within the *FMO3* gene, which results in a reduced or abolished ability of the liver to convert TMA to TMAO (D'Angelo et al., 2013). The excess TMA is excreted via urine, sweat, and breath, producing a pungent odour, and those who suffer from this condition experience significant social and psychological impacts (Ayesh et al., 1993). Additionally, TMA *per se* is a recognised uraemic toxin (Jaworska et al., 2019) and its concentration in serum has also been reported as having a positive association with cardiovascular disease pathogenesis, including foam cell formation associated with atherosclerotic lesions (Bordoni et al., 2020; Koeth et al., 2013; Roncal et al., 2019). Providing TMA-utilising methanogens to individuals with this condition is thus thought to potentially lower gut and consequently serum TMA and TMAO concentrations. In fact, this hypothesis was recently tested by the longitudinal quantification of TMA(O) in a mouse model inoculated with different environmental and host-associated methanogen species (Ramezani et al., 2018). Interestingly, the hydrogenotrophic human methanogen *M. smithii,* along with environmental isolates *M. mazei* and *M. blatticola*, were shown to be more effective at gut colonisation and reducing serum TMAO concentrations than the TMA-utilizing *Methanomassiliicoccus* strains tested in this study and further, colonisation (as measured by RT-qPCR) was inversely proportional to the serum TMAO concentration (Ramezani et al., 2018). Despite these findings, there is currently no evidence that the *M. smithii* strain used in these studies (DSMZ861^T) can use TMA or other methylated amines as a primary substrate for methanogenesis.

1.13 Opportunities for translational research beyond human health

In the preceding sections, I have provided an overview of the emerging and growing interest in the potential roles and opportunities of human gut methanogens in health and disease. This has been driven in large part by the "capture" of methanogen (and archaeal) diversity by the era of metagenomic sequencing and as outlined in the following Chapters, a wealth of genomic information from which to extract new knowledge. However, the interest in gut methanogen diversity and ecology by the agricultural, environmental, and engineering sectors has been longstanding, spurred largely by society's concerns about global change in climate, resource utilisation, as well as energy and environmental sustainability. As such, how might the knowledge base for gut methanogenic archaea generated from human microbiome research be translated and used to address key knowledge gaps for other research challenges, with a focus towards a broader range of societal needs? In the following sections, I will discuss the methanogenic communities of wild and agriculturally relevant domesticated herbivores, and the role they play in the differentiation of methane production by "high-methane" and "low-methane' emitting animals. Additionally, I will also discuss how the study of these communities may be key in the reduction of methane emissions, as well as how they may be utilised to understand ecological effects on wild animal populations.

1.14 Methane emissions from domesticated ruminants and their production systems

The digestive anatomy of ruminant animals is typified by a multi-chambered system comprised of the rumen, reticulum, and omasum that precedes the stomach (abomasum). The rumen-reticulum provides a suitable habitat for microbes from all three Domains of life. This community is characterised by its degradative properties towards plant biomass (including cellulose) and anaerobic fermentation, with 2-12% of digestible energy released as methane (Chaokaur et al., 2015; Johnson & Johnson, 1995). Domesticated species of ruminants used for meat, fibre and milk production are estimated to produce 87-97Tg (CH_4/year) of methane annually (Grossi et al., 2018), which contribute to at least 20% of global methane emissions (Grossi et al., 2018; Johnson & Johnson, 1995; Lassey, 2007; Leahy et al., 2010). Targeting methanogenic archaea in livestock, specifically ruminants, remains a hot topic to reduce agricultural methane emissions. As such, understanding the diversity and nutritional ecology of agriculturally relevant methanogens is essential in efforts to mitigate methane production. A large variety of methanogens have been detected and isolated from agricultural animal samples including species of *Methanobrevibacter, Methanosphaera, Methanomicrobiales,* and *Methanobacterium,* as well as *Methanomassiliicoccales*-associated species. Interestingly, isolation of heterotrophic, acetate utilising *Methanosarcina* are relatively rare (Table 1.2). Studies characterising the microbiome of low methane-emitting and high methane-emitting sheep and beef cattle have shown an increase in *Methanosphaera* species in low methane-emitting

sheep and a relative increase in *Methanobrevibacter* in high methane-emitting sheep (Shi et al., 2014). Similar work by Martínez-Álvaro et al. (2020) showed low-methane emitting bovine to contain a greater diversity methanogens involved in the three methanogenesis pathway, with an increase in *Methanosphaera* and *Methanomethylophilus*, compared to high-methane emitting

Table 1.2. Cultured methanogens isolated from animal hosts, for which genomic data are available. All information was obtained from the NCBI Genome database. * denotes multiple isolated strains for a given species.

	***Methanobrevibacter*-associated**	**Type Strain**	**Host**	**Country**	**Taxon ID**	**RefSeq ID**
	Methanobrevibacter sp. YE315	YE315	*Bos taurus indicus*	Australia	1609968	NZ_CP010834.1
	Methanobrevibacter ruminantium	M1	Bovine	New Zealand	634498	NC_013790.1
	Methanobrevibacter thaueri	DSM 11995	Bovine	Germany	190975	NZ_MZGS00000000.1
	Methanobrevibacter woesei	DSM 11979	Goose	Germany	190976	NZ_MZGU00000000.1
	Methanobrevibacter boviskoreani	JH1	HanWoo steer	South Korea	1214066	NZ_BAGX00000000.2
	Methanobrevibacter arboriphilus	ANOR1	*Homo sapiens*	France	1401244	NZ_CBVX000000000.1
*	*Methanobrevibacter smithii*	PS	*Homo sapiens*	USA	420247	NC_009515.1
	Methanobrevibacter oralis	DSM 7256	*Homo sapiens* (oral)	Germany	66851	NZ_LWMU00000000.1
*	*Methanobrevibacter gottschalkii*	DSM 11978	Horse/Pig	USA	190974	NZ_FOAK00000000.1
	Methanobrevibacter sp. AbM4	AbM4	Ovine	New Zealand	224719	NC_021355.1
	Methanobrevibacter wolinii	SH	Ovine	USA	1410664	NZ_JHWX00000000.1
	Methanobrevibacter sp. 87.7	87.7	Ovine (Lamb)	France	387957	NZ_MRCT00000000.1
	Methanobrevibacter millerae	SM9	*Ovis aries*	New Zealand	230361	NZ_CP011266.1
	Methanobrevibacter olleyae	YLM1	*Ovis aries*	New Zealand	294671	NZ_CP014265.1
	Methanobrevibacter curvatus	DSM 11111	*Reticulitermes flavipes*	Germany	49547	NZ_LWMV00000000.1
	Methanobrevibacter cuticularis	DSM 11139	*Reticulitermes flavipes*	Germany	47311	NZ_LWMW00000000.1
	Methanobrevibacter filiformis	DSM 11501	*Reticulitermes flavipes*	Germany	55758	NZ_LWMT00000000.1
	***Methanomassiliicoccales*-associated**	Type strain	Host	Country	Taxon ID	RefSeq ID
	Candidatus Methanomethylophilus sp. 1R26	1R26	*Bos taurus*	Denmark	1769296	NZ_LOPS00000000.1
	Thermoplasmatales archaeon BRNA1	BRNA1	Bovine	Australia	1054217	NC_020892.1
	Candidatus Methanomassiliicoccus intestinalis	Issoire-Mx1	*Homo sapiens*	France	1295009	NC_021353.1
	Candidatus Methanomethylophilus alvus	Mx1201	*Homo sapiens*	France	1236689	NC_020913.1
	Methanomassiliicoccus luminyensis	B10	*Homo sapiens*	France	1175296	NZ_CAJE00000000.1
	Methanogenic archaeon ISO4-H5	ISO4-H5	*Ovis aries*	New Zealand	1495144	NZ_CP014214.1
	Methanogenic archaeon ISO4-G1	ISO4-G1	*Ovis aries*	New Zealand	1452364	CP013703.1
	Candidatus Methanoplasma termitum	MpT1	Termite hindgut	Germany	1577791	NZ_CP010070.1
	***Methanosphaera*-associated**	Type strain	Host	Country	Taxon ID	RefSeq ID
	Methanosphaera sp. BMS	BMS	*Bos taurus indicus*	Australia	1789762	NZ_CP014213.1
*	*Methanosphaera stadtmanae*	DSM 3091	*Homo sapiens*	USA	339860	NC_007681.1
	Methanosphaera sp. WGK6	WGK6	*Macropus fuliginosus*	Australia	1561964	NZ_JRWK00000000.1
	Methanosphaera cuniculi	1R-7	*Oryctolagus cuniculus*	USA	1077256	NZ_LMVN00000000.1
	***Methanobacterium*-associated**	Type strain	Host	Country	Taxon ID	RefSeq ID
	Methanobacterium formicicum	BRM9	Bos taurus	New Zealand	2162	NZ_CP006933.1
	Methanobacterium subterraneum	DF	Deer	Netherlands	59277	JABBYL010000000
	***Methanosarcina*-associated**	Type Strain	Host	Country	Taxon ID	RefSeq ID
	Methanosarcina barkeri	CM1	Bos taurus	New Zealand	796385	NZ_CP008746.1
	***Methanomicrobium*-associated**	Type strain	Host	Country	Taxon ID	RefSeq ID
	Methanomicrobium mobile	BP	Bovine	USA	694440	NZ_JOMF00000000

bovine, which contained a higher abundance of *Methanobrevibacter.* Additionally, they also show the key explanatory variables between the group to be complex carbohydrate degradation, metabolism of sugars, and metabolism of amino acids, suggesting the interaction of methanogens with their microbial community is also an essential component of methane mitigation.

As such several studies have attempted to define methods for the targeted suppression of methanogen species. One study by Tan et al. (2011) focused on the use of tannins for the suppression of methane production. This study showed a reduction in methane emission, with an increase in the diversity of *Methanomassiliicoccales*, and a relative decrease in the diversity of *Methanobacteriales* and *Methanomicrobiales* (Tan et al., 2011). Similar studies comparing cattle and yaks have shown a different methanogen community and increased diversity in lower methane-emitting yak compared to high methane-emitting cattle (Huang et al., 2012). Despite the characterised changes in the methanogenic communities, little is known about the genetics of these methanogens and how they differentially contribute to the emission of low- and high-methane emitting animals. As such, my work will provide an expansion of available genetic information on the methanogen communities of various methane-emitting animals to provide a comprehensive comparison of factors potentially associated with a high methane producing phenotype.

1.15 Non-ruminant and native Australian herbivores

Non-ruminant herbivores are also dependent on the recruitment and retention of microbes within different segments of their gastrointestinal tract in support of plant biomass conversion to nutrients. These adaptations to herbivory are outlined in detail by Mackie et al. (2000) but in brief detail, these animals may utilise either a sacciform, non-gastric region of the stomach (e.g. the macropodids, such as kangaroos and wallabies) or hindgut (caecum or colon)(White & Mackie, 1997). Duroc, Landrace, Yorkshire (Mao et al., 2011; Mi et al., 2019), Erhualian and Landrace (Zhu et al., 2011) pigs also contain a significant abundance of *Methanobrevibacter*. One study on Canadian pigs showed *Methanoculleus* spp. as additional major contributors to methane emissions through hydrogenotrophic methanogenesis (Barret et al., 2013). Colonic fermenters such as horses have shown an abundance of *Methanobrevibacter* (Fernandes et al., 2014; Lin & Miller, 1998) and *Methanocorpusculum* species (Fernandes et al., 2014; MM et al., 2013). Similarly, white and black rhinoceros are colonised by *Methanobrevibacter* and *Methanocorpusculum*, as well as *Methanosphaera* and *Methanomassiliicoccales*-related species (Gibson et al., 2019; Luo et al., 2013). In fact, *Methanocorpusculum* spp. was the most abundant methanogen in captive white rhinos at ~60% (Luo et al., 2013) and has also been found as the predominant taxon in Japanese thoroughbred horses and ponies (Lwin & Matsui, 2014). Smaller caecal fermenting animals also show methanogen colonization. *Methanobrevibacter*-related species have been isolated from rodents such as the faeces

of rats, with ageing rats showing an increased abundance of total methanogens (Lin & Miller, 1998; Maczulak et al., 1989). Additionally, squirrels also contain *Methanosphaera* (Carey et al., 2013). The caecal contents of rabbits have also shown the presence of *Methanosphaera* and several *Methanobrevibacter* species (Kusar & Avgustin, 2010). Additionally, *Methanosphaera cuniculi* was isolated in pure culture from the intestinal tract of a rabbit, proving genomic insights into the limited number of characterised *Methanosphaera*. The diet of the North American beaver consists entirely of woods, roots and aquatic plants, requiring a syntrophic relationship with fermentative bacteria to aid in digestion (Kohl et al., 2014). Interestingly, *Methanosphaera, Methanobrevibacter,* and *Thermoplasmatales* were detected from the caecum and faeces, but *Methanosphaera*-associated OTUs accounted for more than 99% of archaeal reads (Kohl et al., 2014). Multiple sequences of *Methanosphaera* were detected across all samples, with a single OTU accounting for 85-90% of *Methanosphaera* sequences (Kohl et al., 2014). This shift towards *Methanosphaera* is likely driven by the production of methanol by bacterial fermentation of pectin derived from the plant-rich diet (Pieper et al., 1980; Revilla & González-SanJosé, 1998).

Studies on members of Macropodidae have shown significantly lower relative methane production compared to conventional livestock when provided the same food source (Von Engelhardt et al., 1978). Additionally, Evans et al. (2009) showed methanogens to colonize the macropodid gut at a significantly lower abundance compared to ruminants. Despite the decrease in abundance, the communities of methanogens were determined to be phylogenetically similar between both groups (Evans et al., 2009). Evans (2011) successfully isolated or enrich cultured representatives of *Methanobrevibacter* and *Thermoplasmatales*-affiliated archaea (*Methanomassiliicoccales*-affiliated) from tammar wallabies (*Macropus eugenii*) and Western grey kangaroos (*Macropus fuliginosus*). To further explore these 'low-methane' methanogens, Hoedt et al. (2016) successfully isolated a representative of *Methanosphaera* (sp. WGK6) from the forestomach contents of a western grey kangaroo (*Macropus fuliginosus*). Unlike all other current *Methanosphaera* isolates, *Methanosphaera* sp. WGK6 was shown to readily utilise ethanol and methanol as substrates for methanogenesis. Despite the observed 'low-methane' emissions of macropodids, *Methanosphaera* sp. WGK6 was shown to produce a greater methane output using methanol/ethanol, compared to methanol/H_2 and the human isolate *M. stadtmanae* when grown on methanol/H_2.

Koalas (*Phascolarctos cinereus*) represent a unique marsupial lineage in that their diet is highly specialised and consists entirely of foliage from *Eucalyptus* trees. In addition to being nutritionally lacking, this plant-based diet results in the ingestion of toxic and antimicrobial plant secondary metabolites (Eberhard et al., 1975; Eschler et al., 2000). Thus, the microbiome plays a large part in the maintenance and degradation of these toxic compounds (Osawa et al., 1993; Osawa et al., 1995).

In terms of methanogens, analysis of the koala and southern hairy-nosed wombat (*Lasiorhinus latifrons*) showed that both animals contain *Methanocorpusculum,* with the wombat containing a significantly greater abundance (2.14%) compared to the koala (0.11%)(Shiffman et al., 2017). Given the lower abundance of methanogens in koalas, it has been speculated that one or more *Eucalyptus*-associated metabolites could be used for the targeted reduction of methanogens, and thus methane emissions, from livestock though this is yet to be tested (Cieslak et al., 2013).

As the human population and urban development put pressure on the natural habitat of different wildlife species, the microbiome may give specific insight into at-risk species. Similar conservational attempts are being made with at-risk species such as the Tasmanian devil (Cheng et al., 2015) and black rhinoceros (Gibson et al., 2019). Along these lines, specific archaeal species may be used as biomarkers for affected wildlife species. With the majority of agricultural animals dominated by *Methanobrevibacter* species, a shift towards *Methanobrevibacter* in wild animals could be indicative of negative ecological changes. One study on black rhinoceros showed there to be an increase in the abundance of *Euryarchaeota* in those held in captive (Gibson et al., 2019). Additionally, an increased abundance of *Methanobrevibacter ruminantium* was observed in captive rhinos compared to wild rhinos, which showed a higher relative abundance of *Methanocorpusculum bavaricum* (Gibson et al., 2019). Similarly, captivity was shown to increase the relative abundance of *Methanobrevibacter* and *Methanosphaera* species in red-shanked douc and mantled howling monkeys (Clayton et al., 2016). A larger study on the effects of captivity on the mammalian gut microbiome showed that most of the 41 tested mammalian species were significantly impacted by captivity (McKenzie et al., 2017), likely also causing a relative shift in archaeal communities.

1.16 Summary and research objectives

With the increasing efficiency and cost-effectiveness of molecular sequencing techniques, studies have increasingly shown a diverse ‘archaeomes’ across wild and captive animal species from various habitats. Currently, little is known about how the variations in these archaea, specifically methanogens, effect their host species. With certain human and animal populations being enriched for specific lineages of methanogenic archaea, it is important to understand the resulting effect on the host and the surrounding microbiome. *Methanosphaera* are typically found in as low abundance members in the human gastrointestinal tract but see increased prevalence in individuals with IBD, concurrent with a decreased abundance of *Methanobrevibacter* (Blais Lecours et al., 2014). *Methanocorpusculum* species have only been successfully isolated from environmental samples but have been recorded in high abundance in rhinoceros, horses, ptarmigan, and some chicken populations, where *Methanobrevibacter* species would typically be expected as the dominant genera

(Hou et al., 2016; Luo et al., 2013; Lwin & Matsui, 2014; Salgado-Flores et al., 2019). Indeed, *Methanocorpusculum* has been detected in specific populations of humans, such as a recent study on Fijian individuals (Brito et al., 2019).

My work will focus on these unique archaeal species, including *Methanosphaera* and *Methanocorpusculum*, to determine the potential reasons they are enriched in certain gut environments or in host species, and the resulting metabolic contributions they provide to their surrounding environment. I will do this by isolating key methanogen species of interest from both human and animal digestive tracts and assess their metabolic potential through culture- and genomics-based analyses.

Specifically, my PhD research is presented across three research Chapters.

1. The isolation of human *M. smithii* and *M. stadtmanae* spp. and respective analysis of their tolerance and metabolism of bile salts; a key environmental factor in gastrointestinal health and disease.
2. Recovery of human-associated methanogen MAGs from publicly available metagenomic datasets and comparative genomic analyses to assess their comparative genetic potential with applied host metadata.
3. Isolation and characterisation of *Methanocorpusculum* and other novel methanogen spp. from Australian marsupials through genomic- and culture-based work. As well as a subsequent comparative genomic analysis of MAGs recovered from diverse animal species.

Chapter 2: Genomic- and culture-based analyses of bile salt metabolism by human methanogenic archaea

2.1 Introduction

Though methanogens are consistently observed as members of the healthy human gut microbiota, there is increasing evidence that they may play a more active role in the progression or maintenance of gastrointestinal diseases. Typically, *Methanobrevibacter* species appear as the dominant methanogens observed in healthy individuals, but recent studies have shown a decreased abundance in patients with IBD (Blais Lecours et al., 2014; Ghavami et al., 2018). These patients also displayed an increase in the prevalence of *Methanosphaera* spp. relative to healthy controls, with an additional *Methanosphaera*-specific IgG response observed in patients with IBD where no significant *Methanobrevibacter*-specific response was observed (Blais Lecours et al., 2014). Similarly, Bang et al. (2014) showed whole-cell preparations of *M. stadtmanae* induced a much stronger pro-inflammatory immune response from moDCs than *M. smithii*. Collectively, these studies suggest that *Methanosphaera* spp. may play a more active role in host-microbe interactions than previously considered, and particularly, in patients with IBD.

One key ecological change associated with IBD is a shift in the bile acid pool, with an increase in the relative concentration of primary and sulphated bile acids, as well as an increase in conjugated bile acids compared to healthy individuals (Duboc et al., 2013; Joyce & Gahan, 2017; Staley et al., 2017). As a part of a larger analysis on the genomic adaptations of *M. smithii*, Samuel et al. (2007) showed *M. smithii* to contain a predicted *bsh* gene. Subsequently, the recombinant gene expressed in *E. coli* was shown to efficiently deconjugate both glyco- and tauro-chenodeoxycholic acids (Jones et al., 2008). Additional studies have also identified a predicted *bsh* homolog within the genome of *M. stadtmanae* (Fricke et al., 2006). The identification and characterisation of the methanogen *bsh* genes have largely been performed on a small subset of genomes, providing an important but relatively narrow range of *bsh* genes. To date, no analysis of bile salt tolerance and metabolism has been conducted on axenic cultures of human methanogens *in vitro*, to infer what effects bile acids have on methanogens within the larger gut environment.

Here I show culture- and genomics-based analyses of bile salt metabolism by different strains of human methanogens. To this end, I recovered new strains of *M. smithii* and *M. stadtmanae* from an Australian population. Using these new isolates and type strains, I show the effect of bile salts on the growth kinetics of *M. smithii* and *M. stadtmanae in vitro* and show strain-specific bile acid metabolite profiles through polar metabolomics. Additionally, I perform a comparative analysis of human methanogen *bsh* using isolate genomes and recently recovered MAGs.

2.2 Materials and Methods

2.2.1 Methanogen enrichment and isolation

Stool samples used for methanogen enrichment and isolation were provided by healthy volunteers and collected under the UQ Human Research Ethics Approval (HREC) #2015000775 (*M. smithii* JC9 sample) or healthy adult volunteers recruited for a nutritional trial performed at Monash University MU-HREC CF14/2904 – 2014001593 and UQ-HREC #2015000317 (*M. smithii* PAM sample). Samples were cryopreserved and subsampled as described by Hoedt et al. (2016). Two separate enrichment series were performed: BRN-RF30 medium (enrichment series one) (Hoedt et al., 2016; Joblin et al., 1990) and BRN-RF10 medium (enrichment series two) were prepared in 10 mL Balch tube aliquots, with substrate combination of 250 mM methanol/H_2, 20 mM TMA/H_2, 20 mM sodium acetate/H_2 or CO_2/H_2 (20:80). All gas was pressurised to 150kPa. Each was also supplemented with ampicillin (500 µg/mL) and streptomycin (80 µg/mL) to suppress bacterial growth. ~200 µL of respective faecal samples was inoculated into each tube, which were then incubated at 37°C with 100 rpm rotational agitation. After 7 days, a subsample of each culture's headspace was analysed for methane production using a Shimadzu GC-2014 (Shimadzu, Kyoto, Japan) fitted with a flame ionization detector for CO_2, H_2, and CH_4, as described by Gagen et al. (2014). The cultures were also analysed for auto-fluorescent using a Zeiss AX10 epifluorescence microscope at 420 nm with a cyan (47 HE) filter set. Positive cultures were subjected to a dilution series with additional vancomycin (50 µg/mL) until subsamples tested negative for bacterial contamination by bacteria-specific PCR with 27F/1492R primers (Enticknap et al., 2006). Cultures positive for methane and negative for bacteria were inoculated into "roll tubes" containing either BRN-RF30 agar (strain JC9) or BRN-RF10 agar (strain PAM) with 125 kPa H_2:CO_2 (80:20), as described by Hungate and Macy (1973). After ~14 days of incubation at 37°C, single colonies were picked and grown in their respective medium and substrates. 16S rRNA PCR amplicons produced using archaea-specific 86F/1340R primers (Wright & Pimm, 2003) and sequenced at the Australian Genome Research Facility (AGRF) for initial taxonomic classification. The enrichment and isolation of *M. stadtmanae* PA5 is described by Hoedt et al. (2018)

2.2.2 *M. smithii* JC9 and PAM DNA extraction and genome sequencing

M. smithii JC9 was grown to mid-exponential phase in 100 mL of BRN-RF30 medium with a headspace of H_2:CO_2 (80:20, 150 kPa). The culture was pelleted at 3200 x g for 15 min and genomic DNA (gDNA) extracted as described by Hoedt et al. (2018) using consecutive freeze-thaws for cell lysis. DNA quality was then assessed using gel electrophoresis and the Quantifluor® dsDNA System (Promega), as per the manufacturer's instructions. *M. smithii* PAM was grown to mid-exponential

phase in 10 mL of BRN-RF10 medium with a headspace of H_2:CO_2 (80:20, 150kPa). Biomass was then pelleted at 13000 x g for 10 min and resuspended in 600 µL RBB+C lysis buffer (Yu & Morrison, 2004). 0.4 g of acid-washed, sterile zirconium beads (0.1 mm + 1 mm) was added and the samples homogenized using a Precelley's 24 Tissue Homogeniser at 5000 rpm (3 x 60 s) with a subsequent incubation at 70°C for 15 min. gDNA was then extracted from the cell preparation using the Maxwell SEV Tissue Kit on a Maxwell 16 Research Instrument (Promega, Madison, WI, USA), as per the manufacturer's instructions. The gDNA was then used to produce short read sequence data using the Illumina NextSeq workflow at the Australian Centre for Genomics (ACE, www.ecogenomic.org). Raw sequences were assembled using CLC Genomics Workbench 10 (www.qiagenbioinformatics.com). CheckM (v1.1.2) (Parks et al., 2015) was used to assess the quality of the genome assemblies and GTDB-tk (v1.3.0) was used to assign taxonomy (Chaumeil et al., 2019). Annotation of predicted ORF was conducted using Prokka (v1.12) (Seemann, 2014). Sequencing and assembly of *M. stadtmanae* PA5 is described by Hoedt et al. (2018)

2.2.3 Recovery of *M. smithii* MAGs from publicly available metagenome datasets

Human faecal metagenome datasets produced by Wang et al. (2018) and Franzosa et al. (2019) were used for methanogen MAG recovery. Sequences were downloaded from the NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra) and trimmed using Trimmomatic v0.32 (Bolger et al., 2014) Sequences were assembled using MegaHit (v1.1.1) (Li et al., 2015) and reads were mapped to the assembly BamM (v1.7.3) (https://github.com/Ecogenomics/BamM). Sequences were then binned using the '-superspecific' parameter of Metabat (v0.32.4) (Kang et al., 2015) and quality assessed using CheckM (v1.0.7)(Parks et al., 2015). GTDB-tk (v1.3.0) was used to taxonomically classify archaeal MAGs. Those which were classified as archaeal and had a completeness of ≥90% and contamination ≤5% were used for comparative analyses.

2.2.4 Comparative analysis of *bsh* sequences from human methanogenic archaea

As described by Panigrahi et al. (2014), 189 reference bile salt hydrolase (*bsh*) or penicillin v-acylase (*pva*) were downloaded from the NCBI Refseq database (https://www.ncbi.nlm.nih.gov/refseq/). Methanogen sequences annotated as 'bile salt hydrolase', 'conjugated bile salt hydrolase', or 'choloylglycine hydrolase' were also downloaded from the NCBI Refseq database. Kaptive (v0.5.1) (Wyres et al., 2016) was used to recover additional predicted *bsh* sequences using a custom database of the aforementioned archaeal *bsh* (Table 6.1). Kaptive was used to query the archaeal MAGs produced by Pasolli et al. (2019) and the 48 MAGs that I produced. Recovered *bsh* and reference sequences were aligned ClustalO (v1.2.0) (Sievers et al., 2011), phylogeny inferred using MEGA-X (Kumar et al., 2018) and visualised using the Interactive Tree of Life (iTOL) (https://itol.embl.de).

Recovered sequences were also translated to amino acid sequences using EMBOSS Transeq (Madeira et al., 2019; Rice et al., 2000), and phylogenetically analysed as above. Representative *Bsh* alignments were conducted using ClustalO and visualized using UGENE (Okonechnikov et al., 2012). Phylogeny of methanogen MAGs was performed using GTDB-tk (v1.3.0), with marker gene files analysed using FastTree (v2.1.10) (Price et al., 2010) and visualised using iTOL.

2.2.5 *M. smithii* and *M. stadtmanae* growth in the presence of bile salts

Growth studies were conducted using the type strains *M. smithii* PS^T (DSMZ 861, ATCC 35061, JCM 30028) and *M. stadtmanae* MCB-3^T (DSMZ 3091, ATCC 43021, JCM 11832), as well as Australian isolates *M. smithii* JC9 and *M. stadtmanae* PA5 (Hoedt et al., 2018). The growth studies were conducted in BRN-RF10 medium, modified from Joblin et al. (1990), with a headspace of H_2:CO_2 (80:20) pressurised to 150 kPa. Oxoid bile salts (LP0055; Oxoid) were dissolved in anoxic (N_2 gassed) Milli-Q H_2O at 50% (w/v) and filtered using a 25 mm syringe filter (0.2 µm; Corning Inc.). The stock solution was then diluted so that cultures could be supplemented with 100 µL for a final concentration of 1.0, 0.5, 0.25, and 0.1% (w/v). Triplicate cultures were inoculated in biological duplicate with 200 µL of respective mid-exponential phase (OD_{600}=0.2 for *M. smithii* and 1.0 for *M. stadtmanae*) cultures. The cultures were then incubated at 37°C with 100 rpm of rotational agitation and growth measured by OD_{600} using a Balch-tube modified spectrophotometer. Measurements were taken every 2 h where possible and graphed using GraphPad7. Linear regression of the respective exponential phase of growth was used to determine growth rate and maximum yield was determined by the point of maximal OD_{600}. Statistics were performed using Unpaired Student's T-test within GraphPad7 for comparative growth rates and maximum yield, with a P-value of <0.05 being statistically significant.

2.2.6 Total RNA extraction and qRT-PCR analysis of *bsh* expression

1.5 mL aliquots of mid-exponential cultures described in 2.2.5 were aseptically sampled from each culture. Each aliquot was centrifuged at 15,000 x g for 5 min. The supernatant was transferred to a sterile Eppendorf tube and stored at -80°C. The cell pellet was resuspended in 500 µL of RNALater and stored at 4°C until processed further. Total RNA from each RNALater-treated sample was performed as described by Hoedt et al. (2016). In brief, RNALater-treated samples were diluted with 1 mL of 100 mM TE solution (all solutions made using DEPC-treated dH_2O). Diluted samples were then combined with acid-washed zirconium beads (2 mm + 0.2 mm), 300 µL 100 mM TE, 100 µL 10% SDS and 400 µL phenol: chloroform (pH 4.3). Samples were homogenised using a Precelley's 24 Tissue Homogenizer. Total RNA was then extracted from homogenised samples using a Qiagen RNeasy Mini Kit, as per the manufacturer's instructions. Extracts were tested for gDNA contamination by archaeal specific (86F/1340R) PCR, as per Hoedt et al. (2016). Positive samples

were treated with TURBO DNA-*free*™ Kit (Invitrogen) to remove gDNA. Once all samples were free of gDNA (Figure 6.2), the RNA was quality assessed and quantified using a Bioanalyser RNA Nano Chip (Agilent Technologies), as per the manufacturer's instructions, and samples with a RIN score >9 being retained for further analyses (Table 6.4). Samples were reverse transcribed using SuperScript III First-Strand Synthesis System (Invitrogen), as described by Hoedt et al. (2016). NCBI primer design (Ye et al., 2012) was used to design primers specific to the *bsh* of *M. smithii* PS and *M. stadtmanae* DSMZ3091, respectively. Optimum melting temperatures were determined using standard gradient PCR. 100 ng of each cDNA preparation of *M. stadtmanae* DSMZ3091, *M. stadtmanae* PA5, *M. smithii* PS, and *M. smithii* JC9 was amplified using the respective *bsh* primer sets using the standard Native Taq PCR protocol and visualised by gel electrophoresis. An RNA sample without reverse transcription was used as the negative control.

PCR efficiencies (*E*) were calculated for 958A/1100Ar archaeal 16S rRNA primer (DeLong, 1992; Whitford et al., 2001) and the *Methanosphaera bsh* (this study) sets as per Hoedt et al. (2016). Biological duplicates and technical triplicate samples were used at concentrations of 10, 2, 0.4, and 0.08 ng. As per the method described by Rasmussen (2001), primer efficiencies were determined to be 100% for the *Methanosphaera bsh* primers and 108% for the 958A/1100Ar primers. Fold-change in *bsh* expression was determined as per Pfaffl (2001).

$$bsh \text{ transcript fold change} = \frac{E_{\text{Target}}[Ct_{\text{Target}(control-treatment)}]}{E_{\text{Ref}}[Ct_{\text{Ref}(control-treatment)}]}$$

Here, the target represented by the *bsh* transcript and was normalised to the reference 16S rRNA amplicons. The control group was represented by cDNA samples from *M. stadtmanae* DSMZ3091 and *M. stadtmanae* PA5 cultures with no added bile salts. Treatment represents samples from cultures supplemented with 1.0% bile salt.

2.2.7 Confirmation of the predicted nucleotide insertion in the *bsh* of *M. stadtmanae* PA5

Bsh-specific primers were designed for *M. stadtmanae* and *M. smithii* using NCBI Primer Design (Ye et al., 2012). *M. stadtmanae* DSMZ3091, *M. stadtmanae* PA5, *M. smithii* PS, and *M. smithii* JC9 cDNA samples (as per Section 2.2.6) were amplified using the respective *bsh*-specific primers with MangoMix™, as per the manufacturer's instructions. RNA without reverse transcription was used as the negative control, and respective gDNA was used as the positive control. Amplicons were visualised by 1.5% agarose gel electrophoresis at 100 V for four hours.

Specific primers were designed using NCBI Primer Design for the predicted region of truncation of the *M. stadtmanae* PA5 *bsh*. 100 ng of *M. stadtmanae* DSMZ3091 and PA5 cDNA from 0.0% and 1.0% supplemented bile salt cultures was amplified using MangoMix™, as per the manufacturer's instructions. PCR-grade dH_2O and RNA without reverse transcription were used as negative controls. As above, amplicons were visualised using agarose gel electrophoresis. PCR conditions were optimised for each primer set using gradient PCR. Primer sequences and amplification temperatures for all primer sets are available in the appendix (Table 6.3). Standard PCR cycling conditions were an initial denaturation of 2 min at 94°C, followed by 40 cycles of denaturation at 94°C, annealing for 30 s and extension at 72°C for 30 s, with a final 7 min at 72°C. PCR amplicons from *M. stadtmanae* DSMZ3091 and PA5 grown with 1.0% bile salt and gDNA controls were purified using AMPure XP beads, per the manufacturer's instructions. 75 ng of each purified DNA was sequenced at the Australian Genome Research Facility (https://www.agrf.org.au/) by Sanger sequencing. Sequences were subsequently visualised using ChromasLite v2.1.1 (https://technelysium.com.au/wp/chromas/) and trimmed to remove erroneous nucleotide sequences. Final sequences were aligned using MUSCLE in MEGA-X and visualised using UGENE v1.29.0 (ugene.net).

2.2.8 Analysis of primary bile acid deconjugation in culture supernatants by LC-QToF

Aliquots from exponential cultures grown with supplemented bile salt were sampled, as described in Section 2.2.6. Samples were taken in duplicate from *M. smithii* PS, *M. smithii* JC9, *M. stadtmanae* DSMZ3091, and *M. stadtmanae* PA5 cultures supplemented with 0, 0.1, 0.25, 0.5, and 1.0% bile acid. Samples were held at -80°C until processed at Metabolomics Australia (https://www.metabolomics.net.au/). Polar metabolites within each sample were analysed using an Agilent 6545A LC-QToF platform. 13C6-Sorbitol, 13C5,15N1-Valine, and 13C6-Leucine were used as internal standards. 5 µl and 10 µl of randomised samples were aliquoted into microcentrifuge tubes, with the addition of ten times the volume (195 µL/190 µL) of ice-cold extraction/lysis solution (Acetonitrile/Methanol/Water (40:40:20, v/v/v) containing a mixture of 3 internal standards at 4 µM conc.). Samples were vortexed and sonicated in an ice bath for 5 min, with subsequent incubation at 4°C for 10 min on a thermomixer. Centrifugation was used to removed debris from the samples and the lysate was subsequently transferred to HPLC vials for analyses. Pooled biological quality controls (PBQCs) were created by combining 30 µL of each sample. Polar metabolite capture using 7 µL for 10 µL extraction and bile acid capture was performed using 3 µL injection for 5 µL extraction.

The Metabolomics data were statistically analysed and visualised using MetaboAnalyst (https://www.metaboanalyst.ca/). The data was log-transformed and median normalised. The data

was analysed using Principal Component Analysis (PCA), Pattern Search, Hierarchical Heat mapping, and One-way Analysis of Variance (ANOVA).

2.3 Results

2.3.1 Isolation and characterisation of *M. smithii* and *M. stadtmanae* from healthy Australian subjects

Healthy human faecal samples were chosen depending on a high relative abundance of *Methanosphaera* and *Methanobrevibacter*, as identified by previous 16S rRNA sequencing (data not shown). After incubation at 37°C for 7 days, most cultures showed methane production. Epifluorescence microscopy of methane-producing cultures showed short auto-fluorescent bacilli in all cultures, with cultures supplemented with methanol also containing auto-fluorescent diplococci. Given that the auto-fluorescent bacilli likely represented *Methanobrevibacter* in all cultures, the cultures supplemented with acetate and methanol were chosen for further enrichment from the first series of enrichments and only the acetate-supplemented culture from the second series of enrichments. Archaea-specific amplicon sequencing and subsequent taxonomic classification showed both acetate enrichments to be *M. smithii* and the methanol enrichment culture to be *M. stadtmanae*.

The roll tubes of selective agar medium produced colonies after ~14 days, with the acetate-containing cultures producing off-white, circular colonies of ~2 mm diameter. After ~5 weeks, the methanol enrichment produced a homogenous collection of light brown, circular colonies of 1-2 mm diameter. Colonies from each were then picked and grown in liquid culture. As shown in Figure 2.1, both the *M. smithii* isolates presented as 1-2 µm bacilli, often in short chains. Comparatively, *M. stadtmanae* PA5 showed 0.5-1.5 µm diplococci. All three isolates stained Gram-positive and were auto-fluorescent (Figure 2.1). This morphology is consistent with the descriptions for the *M. smithii* and *M. stadtmanae* type strains. The *M. smithii* isolates from the acetate enrichments were given the strain designations JC9 and PAM, and the *M. stadtmanae* isolate was given the strain designation PA5.

The genome assemblies produced for the three strains showed a high level of completeness, with the *M. smithii* isolates being 100% complete and *M. stadtmanae* being 97.6% complete (Table 2.1). All genomes showed no detectable contamination or strain heterogeneity (Table 2.1). As expected, the genomic content of the new isolates showed a high degree of similarity to their respective type strains, although the genome size of the isolates was slightly smaller, in particular for *M. smithii* JC9 which contained 101 fewer genes than *M. smithii* PS^T (Table 2.1). The alignment of the genomes further confirmed the high degree of synteny between the strains (Figure 2.2). Interestingly, the tRNA gene for tryptophan was not detected in any genome, including the type strains.

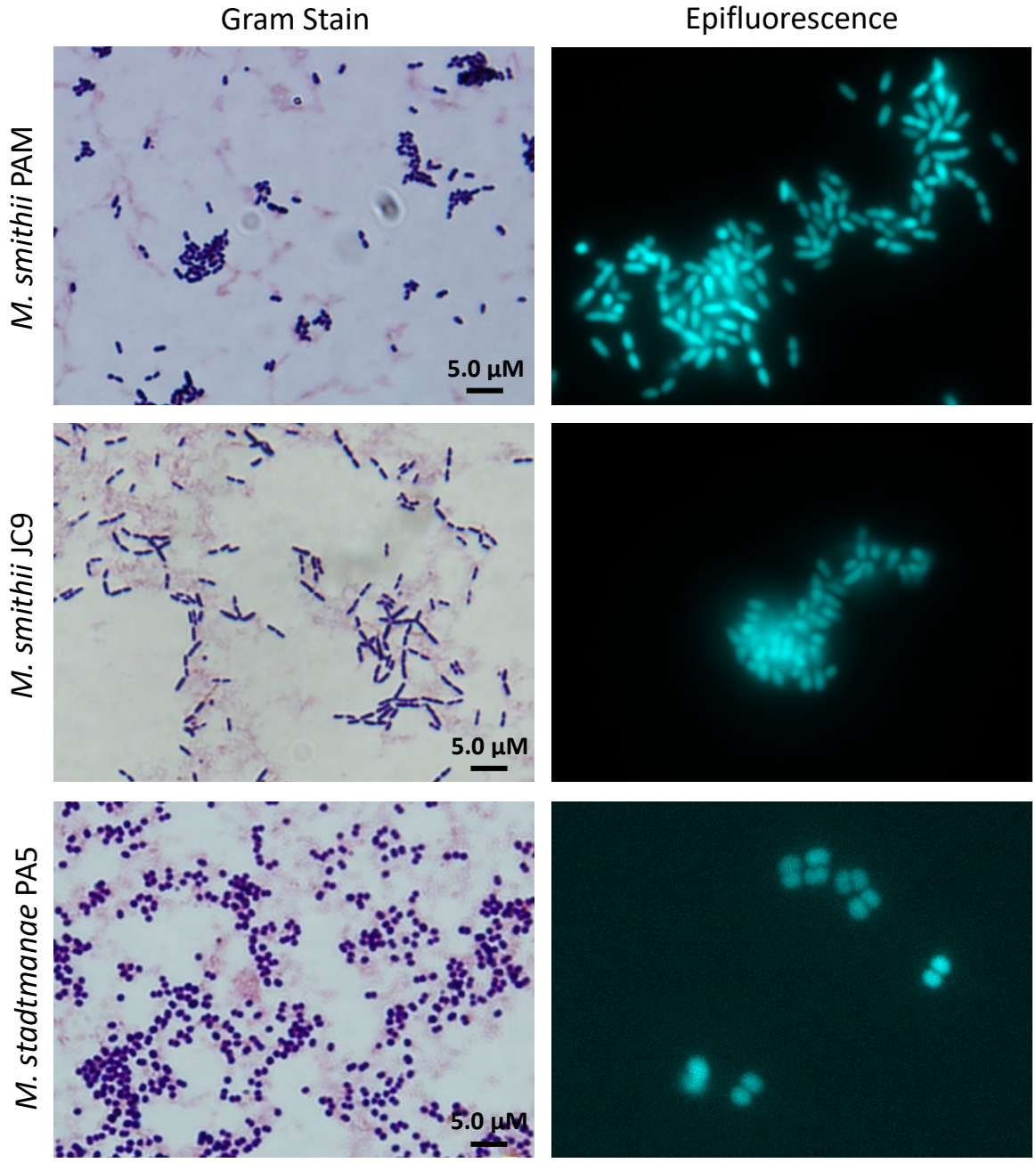


Figure 2.1. Gram stain and epifluorescence micrographs of Australian *M. smithii* and *M. stadtmanae* isolates. Cultures were Gram stained using standard protocols. Epifluorescence micrographs were taken from mid-exponential phase cultures at 420 nm with a cyan (47 HE) filter set. All cultures were auto-fluorescent, with *M. smithii* being short bacilli and *M. stadtmanae* as diplococci.

Table 2.1. Basic genome details of Australian *M. stadtmanae* and *M. smithii* isolates with respective type strains. Genome quality and details were determined using CheckM (v1.1.2). tRNA and rRNA gene counts were annotated using Prokka (v1.12). Taxonomic classification was determined using GTDB-tk (v1.3.0). The genome assemblies of the new isolates were high quality and showed similar genomic content to their respective type strains.

	***M. smithii* PS^T**	***M. smithii* JC9**	***M. smithii* PAM**	***M. stadtmanae* DSMZ3091^T**	***M. stadtmanae* PA5**
Completeness	100	100	100	97.6	97.6
Contamination	0	0	0	0	0
Strain heterogeneity	0	0	0	0	0
No. tRNAs (/20)	19	19	19	19	19
5S rRNA	0	0	0	4	2
16S rRNA	1	1	1	4	1
23S rRNA	1	1	1	4	1
Genome size	1853160	1752600	1771615	1767402	1701735
Contigs	1	26	55	1	131
N50 (contigs)	1853160	107225	131804	1767402	31755
Predicted genes	1782	1681	1772	1536	1546
GTDB classification	s__Methanobrevibacter_A smithii	s__Methanobrevibacter_A smithii	s__Methanobrevibacter_A smithii	s__Methanosphaera stadtmanae	s__Methanosphaera stadtmanae
Geography	USA	Australia	Australia	USA	Australia

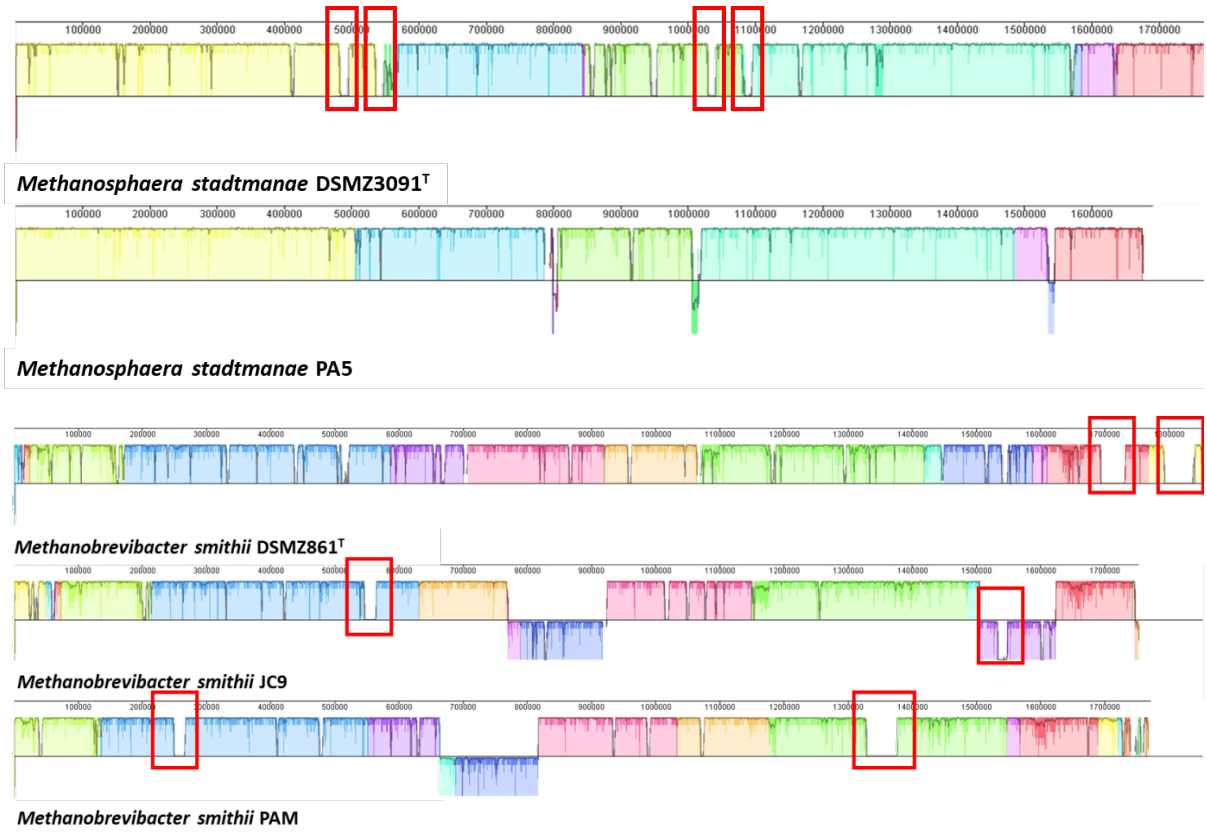


Figure 2.2. ProgressiveMauve alignment of Australian *M. smithii* and *M. stadtmanae* isolates with respective type strain. The genomes assemblies were aligned using progressiveMauve within the MAUVE software package. Contigs of *M. stadtmanae* PA5 were reordered with *M. stadtmanae* DSMZ3091 as the reference. The contigs of the *M. smithii* isolates were reordered with *M. smithii* PS as the reference. All three genomes show a high level of synteny, with only a small number of unique genomic regions found in each respective genome, as outlined by the red boxes.

2.3.2 Effect of bile salts on the growth of human *M. smithii* and *M. stadtmanae*

Human strains of *M. smithii* and *M. stadtmanae* were cultured with Oxoid bile salt supplementation to assess the effect of bile salts on their growth and metabolism (Figure 2.3). For the *M. stadtmanae*, *M. stadtmanae* DSMZ3091 showed a decreasing growth rate with increasing bile salt concentration, with a significant reduction observed in cultures containing 0.5 and 1.0% (Figure 2.3A). Additionally, the higher concentrations also increased the lag time of exponential growth, with consistent maximum yield (Figure 2.3A). Comparatively, *M. stadtmanae* PA5 showed no significant change in growth rate, yield, or lag time at any tested bile salt concentration (Figure 2.3B). Both *M. smithii* strains showed no significant change in growth rate, but the maximum yield was significantly reduced at 0.5 and 1% bile salt relative to cultures containing no additional bile salt (Figure 2.3C-D). Interestingly, this not only shows a genus-level differentiation in bile salt tolerance between the *Methanobrevibacter* and *Methanosphaera* but also strain-level variation within *M. stadtmanae*. Given this variation in bile salt tolerance, supernatant samples were analysed for polar metabolites to assess the presence of bile acids and their associated metabolites.

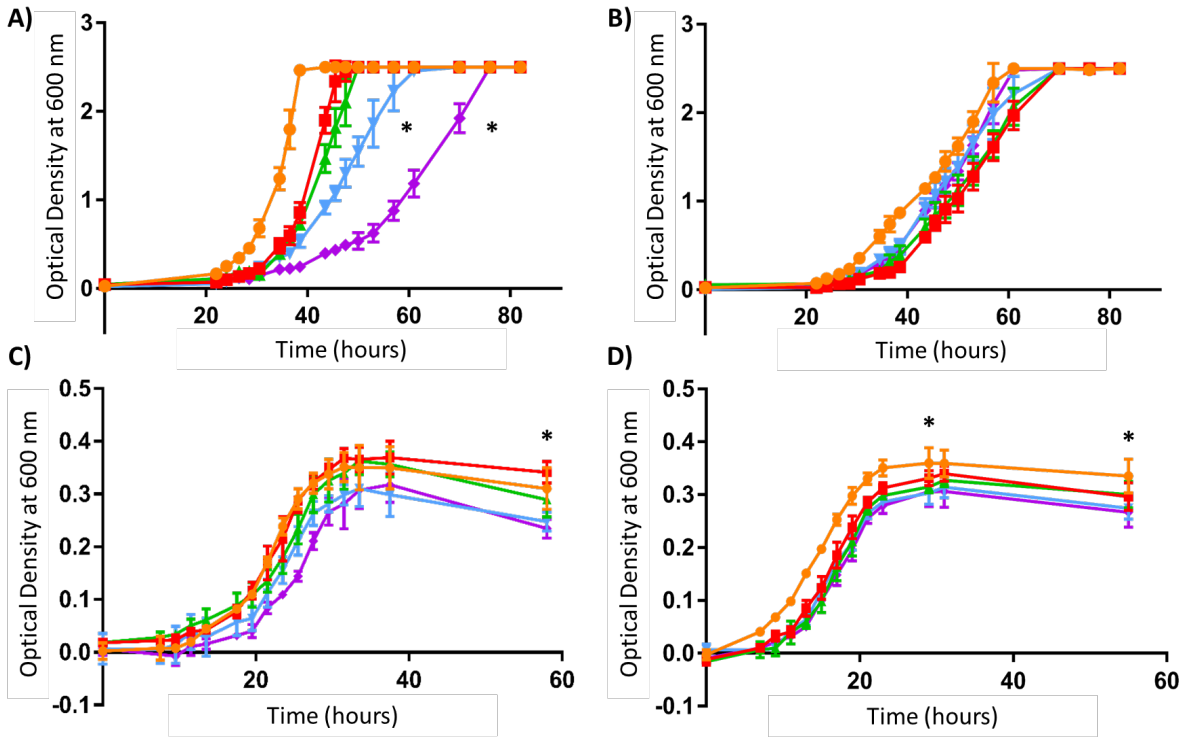
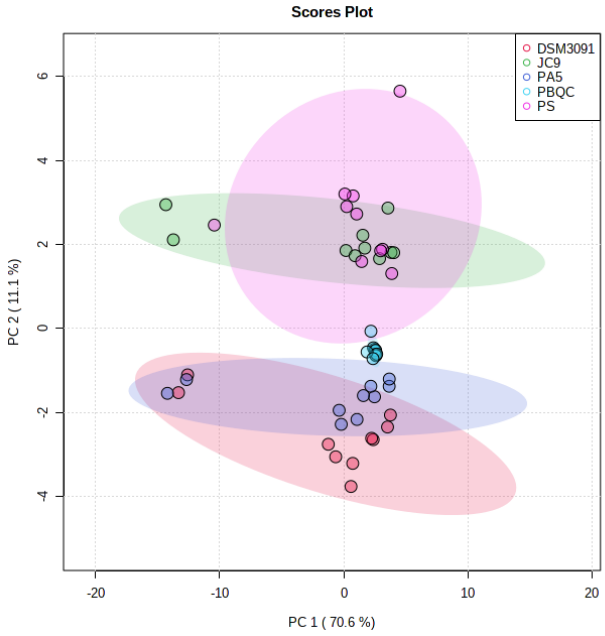


Figure 2.3. Growth kinetics of *M. stadtmanae* DSMZ3091 (A), *M. stadtmanae* PA5 (B), *M. smithii* PS (C) and *M. smithii* JC9 (D) in response to increasing concentrations of bile salts. Concentrations of added bile salts are denoted by line colour and symbol (● 0.0%, ■ 0.1%, ▲ 0.25%, ▼ 0.5%, ♦ 1.0% bile salt [w/v]). Each point represents the mean OD_{600} (± SD) of triplicate cultures in biological duplicate. * represents a statistically significant (p value <0.05) difference in growth rate (A) and maximum yield (C-D).

2.3.3 Detection of polar metabolites in methanogen culture medium by LC-QToF

Supernatant samples from cultures described in Section 2.3.2 were analysed for the presence of bile acid metabolites. 13C5, 15N1 Valine, 13C6-Leucine, and 13C6-Sorbitol were used as internal standards and showed acceptable variance across all samples (Figure 2.4). Endogenous metabolites within the PBQC's showed CV <20% demonstrating the satisfactory performance of the instrument. Total metabolites of each strain were displayed by principal component analysis (PCA) with the PBQC's in Figure 2.4. The PCA showed a tight clustering of the PBQCs between the individual samples, showing sufficient quality of the analysis.

The initial analysis of the samples detected a total of 170 polar metabolites (Table 6.2). Of the detected metabolites, 13 out of a possible 16 bile salts were detected, with sulfolithocholic acid, allocholic acid, and 3b-hydroxy-5-cholenoic acid not detected. Nutriacholic acid and 7a-Hydroxy-3-oxo-5b-cholanoic acid were also determined to be synonymous identifiers and contained identical concentration values, thus the latter of the two was removed from further analyses.



	CV %	
ISTD's	PBQC	Samples
*13C5,15N1-Valine	4.68	14.72
*13C6-Leucine	1.85	14.88
*13C6-Sorbitol	36.24	39.4

Polar metabolites	
CV %	PBQC
< 10 %	91
10 to 20 %	37
20 to 30 %	11
> 30 %	31
Total	170

Figure 2.4. Principal component analysis (PCA) of culture supernatant metabolite profiles with pooled biological quality controls. The polar metabolites of methanogen culture supernatants were analysed by LC-QToF. The data was median normalised and log transformed for visualisation. 170 total polar metabolites were detected in culture supernatants. The metabolite profiles separate according to species, with the PBQCs clustered closely together between all samples. Two of three internal standards variance performed as expected demonstrating sufficient quality of the analyses.

Bile acids were additionally analysed in uninoculated BRN-RF10 medium to ensure the Oxoid bile salts were providing sufficient conjugated bile salts and for relative comparison to spent culture supernatants. Uninoculated BRN-RF10 medium contained trace concentrations of taurocholic acid, tauroursodeoxycholic acid, 3-oxocholic acid, glycocholic acid, and taurodeoxycholic acid. Interestingly, the uninoculated medium also contained a high concentration of ursodeoxycholic acid, likely provided by the rumen fluid component of the medium (Figure 2.5). As can be seen in Figure 2.5, taurocholic acid, glycodeoxycholic acid tauroursodeoxycholic acid, glycocholic acid, and taurodeoxycholic acid increased with increasing supplementation of bile salt, with ursodeoxycholic acid remaining the most abundant bile acid at all concentrations. The non-conjugated cholic acid, nutriacholic acid, deoxycholic acid, 3-oxocholic acid, hyodeoxycholic acid, and lithocholic acid also increased with bile salt supplementation but were lower than the conjugated bile salts. This confirmed

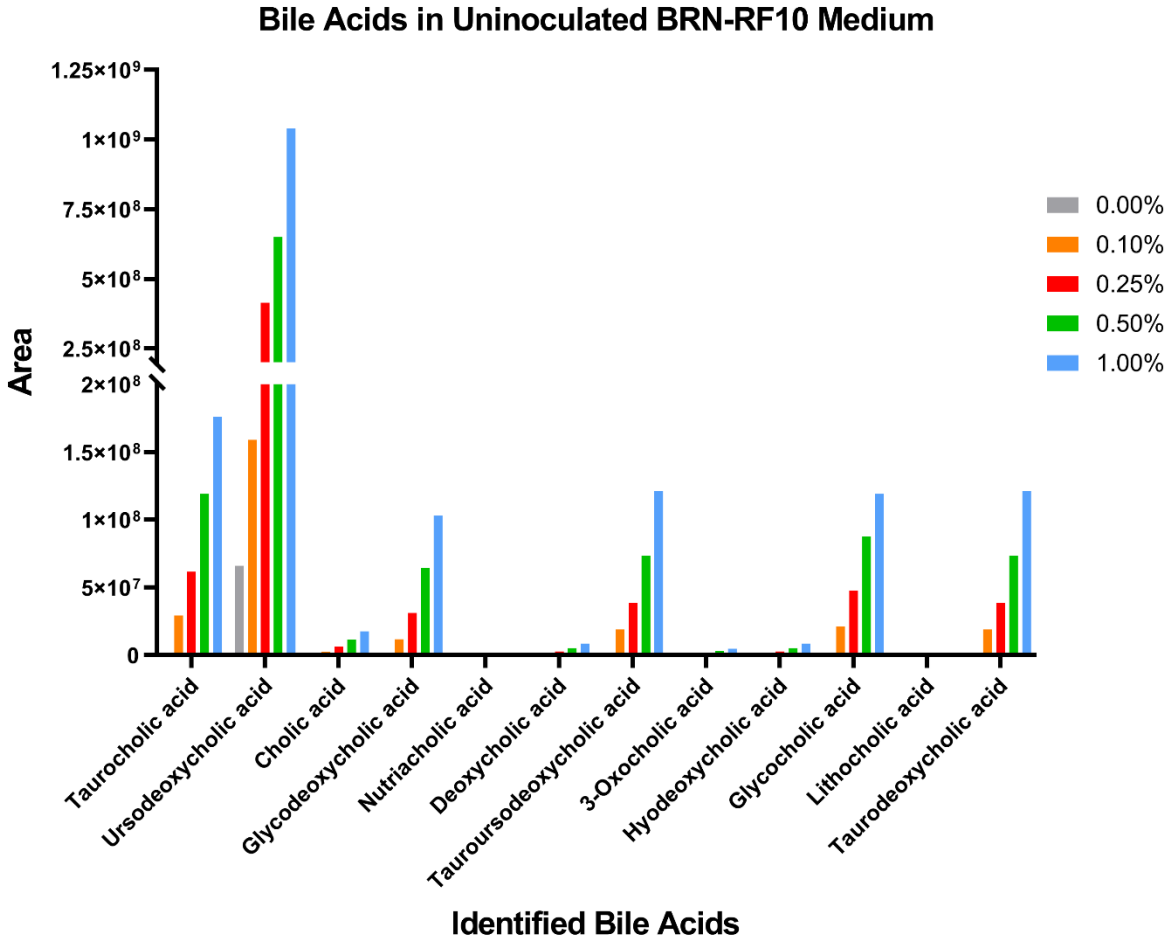


Figure 2.5. Concentration of bile salts detected in BRN-RF10 medium preparations with supplemented Oxoid bile salts. Bile salts represent only those detected by LC-QToF as per the method. Of the bile acids that were detectable using LC-QToF as per the method, allocholic acid, 3b-hydroxy-5-cholenoic acid and sulfolithocholic acid were not detected.

that the Oxoid bile salt preparation provided sufficient bile acids to assess the capability of *M. smithii* and *M. stadtmanae* to deconjugate bile salts *in vitro*.

2.3.4 Metabolite profiles of *M. smithii* and *M. stadtmanae* grown in the presence of bile acid

Bile acid metabolite profiles of the *M. smithii* and *M. stadtmanae* strains showed samples from *M. stadtmanae* PA5 and the *M. smithii* strains to cluster tightly together by PCA, with *M. stadtmanae* DSMZ3091 clustering separately (Figure 2.6A-B). Visualisation of the total 170 polar metabolites by PCA showed a clear differentiation between the *M. smithii* and *M. stadtmanae* strains, with PS and JC9, and DSMZ3091 and PA5 clustering together, respectively (Figure 2.6C-D). When samples without bile acid supplementation were removed, the remaining samples again clustered according to species with reduced variability between samples (Figure 2.6B,D). Analysis by ANOVA again showed significant differences for *M. stadtmanae* DSMZ3091 compared to the other strains (Figure 2.7). Indeed, *M. stadtmanae* DSMZ3091 showed a significant difference in cholic acid (P=4.0889E-8), glycocholic acid (P=7.0516E-21), hyodeoxycholic acid (P=5.7695E-10), deoxycholic acid (P=5.2127E-10), glycodeoxycholic acid (P=4.2836E-21), taurodeoxycholic acid (P=5.0374E-4), and tauroursodeoxycholic acid (P=5.03E-4). An increase in the concentration of the non-conjugated cholic acid and deoxycholic acid, with a concurrent decrease in both glyco- and tauro-conjugated bile acid suggests that *M. stadtmanae* DSMZ3091 can efficiently deconjugate cholic and deoxycholic acid conjugates. Additionally, the reduction of tauroursodeoxycholic acid also suggests it is deconjugated, though there was no significant increase in the concentration of ursodeoxycholic acid. Interestingly, an increase was also observed in the secondary bile acid hyodeoxycholic acid, suggest DSMZ3091 also contains additional capabilities to metabolise primary bile salts. When the uninoculated medium metabolite values were not removed from the culture medium metabolites, nutriacholic, 3-oxo-cholic, and lithocholic also showed additional significant variance between the *M. smithii* and *M. stadtmanae* strains (Figure 6.1). Again, DSMZ3091 showed a significant reduction in the concentration of 3-oxocholic acid (P= 2.7574E-4) but *M. smithii* JC9 showed a higher concentration of nutriacholic acid (P=0.026133) and *M. stadtmanae* PA5 showed a relative reduction in lithocholic acid (P=0.009614)(Figure 6.1).

Given the unusual profile found for *M. stadtmanae* DSMZ3091, a second analysis excluding these samples was performed. The clustering of the three strains remained (Figure 2.8) with one sample (Sample 33: *M. smithii* JC9 replicate 1, with 0.25% bile salts) found to be an outlier. When sample 33 was included, only glycodeoxycholic acid was significantly greater in the *M. smithii* samples (P=0.000148) (Figure 2.8). Glycodeoxycholic acid concentrations were significantly greater in the *M.*

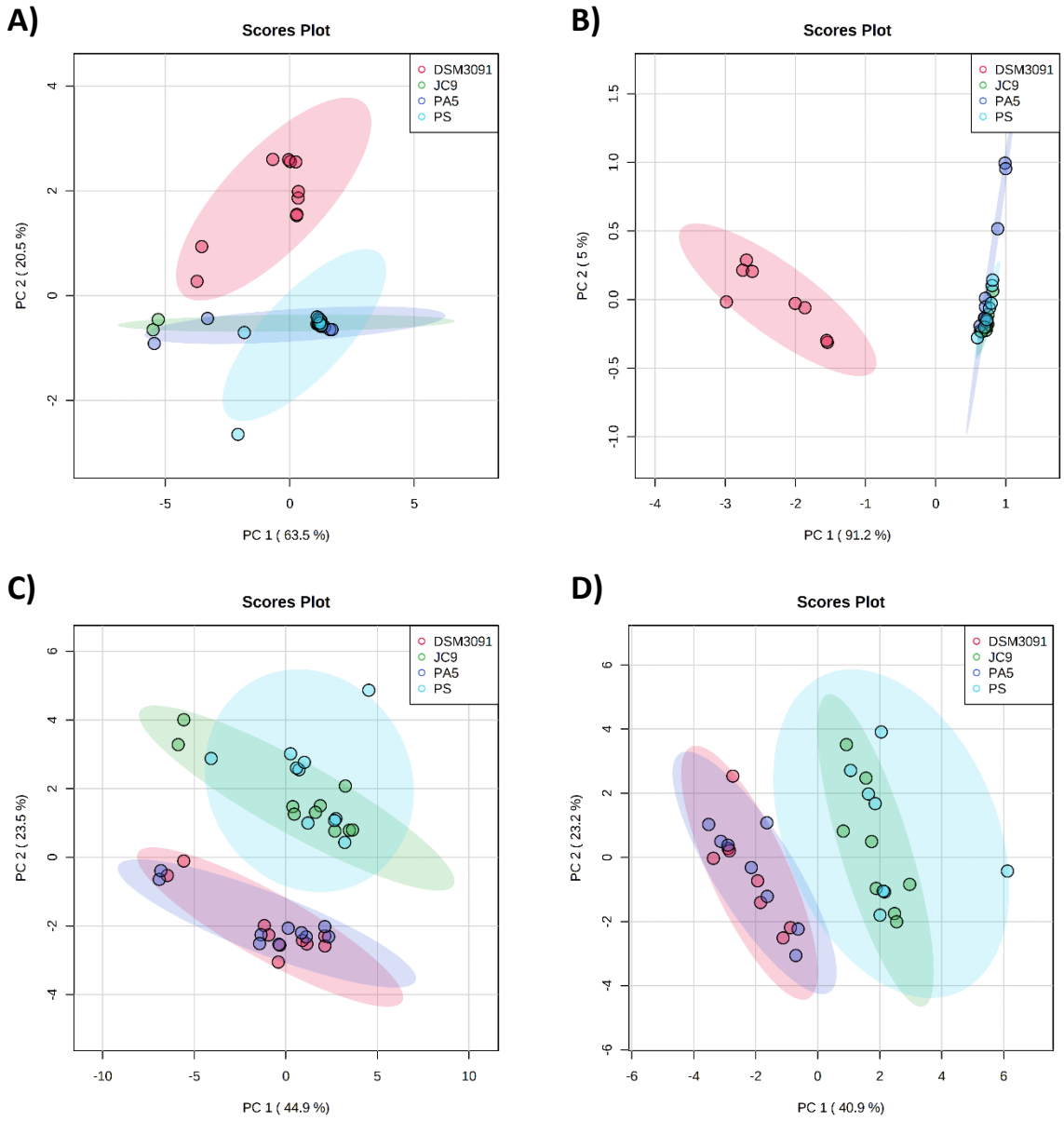


Figure 2.6. Principle component analyses (PCA) of culture supernatant polar metabolite profiles. A) represents PCA of bile acid metabolites identified in all samples, where **B)** represents PCA of bile acid metabolites when samples with 0% supplemented bile salts were removed. **C)** represents PCA of all detected metabolites from all samples, where **D)** represents PCA of all detected metabolites when samples with 0% supplemented bile salts were removed. **C)** and **D)** show the samples separate according to species when the complete metabolite profiles are used. However, DSMZ3091 clusters separately when using the bile acid metabolite profiles, as shown by **A)** and **B)**.

smithii samples (P=0.00038) when sample 33 was excluded in the sub-group analysis but hyodeoxycholic acid was also significantly increased in *M. stadtmanae* PA5 samples relative to the *M. smithii* (P=0.003521; FDR=0.024644) (Figure 2.8).

Along with the unconjugated bile acids, taurine and glycine are liberated as a part of the deconjugation reaction. As such, taurine and glycine concentrations were also analysed across the different strains. As shown in Figure 2.9, taurine showed a strong positive correlation with non-conjugated cholic acid, hyodeoxycholic acid, and deoxycholic acid. Strong negative associations were observed with tauro- and glyco-conjugated bile acids, showing not only tauro-conjugates are metabolised to produce taurine but glyco-conjugated bile acids are similarly being metabolised in these cultures. Interestingly, *M. stadtmanae* DSMZ3091 showed a higher concentration of taurine compared to the other strains,

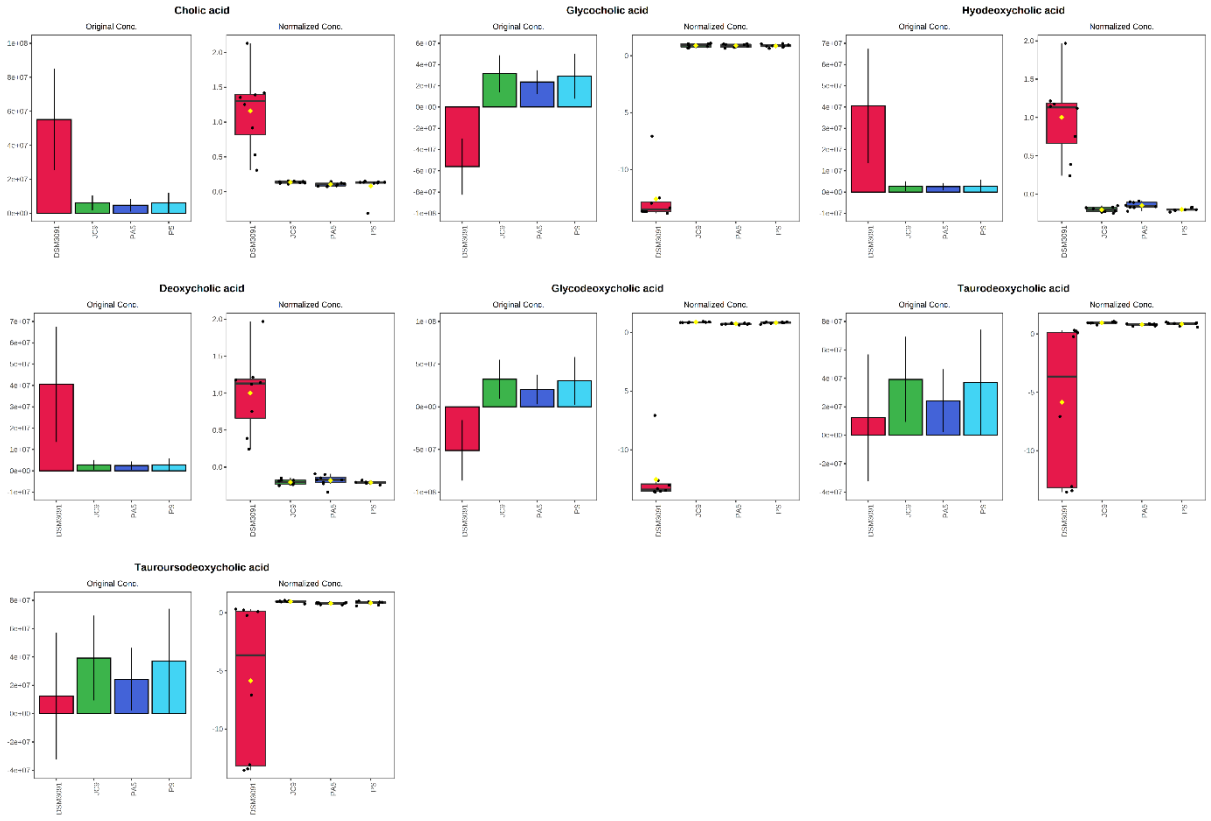


Figure 2.7. Statistically significant bile acid metabolites detected in culture supernatants of *M. smithii* and *M. stadtmanae*. ANOVA was used to compare the bile acid metabolite profiles between each of the strains. Metabolite concentrations of uninoculated BRN-RF10 medium with respective bile salt supplementation were subtracted from the spent culture metabolite concentrations. Culture samples without bile salt supplementation were excluded from this analysis. P values for respective bile acids: cholic acid - 4.0889E-8, glycocholic acid - 7.0516E-21, hyodeoxycholic acid - 5.7695E-10, deoxycholic acid - 5.2127E-10, glycodeoxycholic acid - 4.2836E-21, taurodeoxycholic acid - 5.0374E-4, and tauroursodeoxycholic acid - 5.03E-4.

suggesting these correlations may be largely attributed to this strain, although this difference was not statistically significant. Conversely, glycine was positively correlated with both the tauro- and glyco-

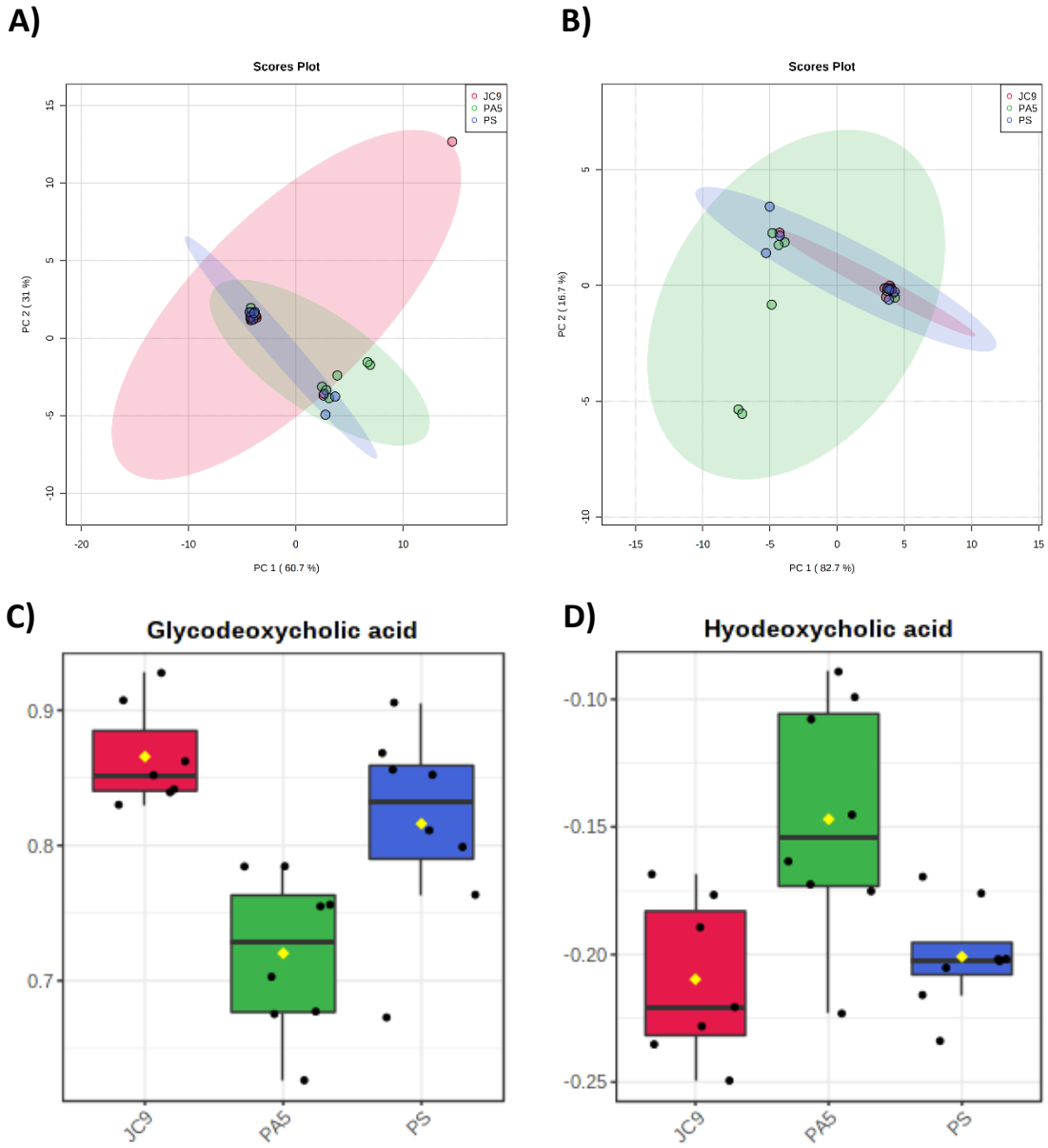


Figure 2.8. Principle component analysis (PCA) and the concentrations of statistically significant bile acid metabolites for culture supernatants. Samples without bile salt supplementation were removed from the analysis. Uninoculated medium concentrations were removed from respective spent culture supernatants. **A)** represents *M. smithii* PS, JC9 and *M. stadtmanae* PA5 samples. **B)** represents when *M. smithii* PS sample 33 (0.25%, replicate 1) was removed. When sample 33 was removed, glycodeoxycholic acid was significantly greater in the *M. smithii* samples **(C)** (P=0.00038; FDR=0.005318) and hyodeoxycholic acid **(D)** was significantly increased in *M. stadtmanae* samples (P=0.003521; FDR=0.024644). When sample 33 was included, only glycodeoxycholic acid was significantly greater in the *M. smithii* samples (P=0.000148; FDR=0.002067).

conjugated bile acids, and negatively correlated with primary bile acids. This is counter to that of taurine and to what would be expected if the glycine produced from these reactions was detected. Although, the free glycine being produced may have been taken up and utilised by the *M. stadtmanae* and *M. smithii* strains, thus resulting in a lower detected concentration and the correlations are affected as a result.

Analysis by hierarchical clustering again showed the expected separation of *M. stadtmanae*. As per Figure 2.8, sample 33 was again shown to be an outlier but grouped with the *M. smithii* and *M. stadtmanae* PA5 samples. For *M. stadtmanae* DSMZ3091, there was a significant reduction glycodeoxycholic and glycocholic acid across all test bile acid concentrations (Figure 2.10). Comparatively, tauroursodeoxycholic acid and taurodeoxycholic acid were only efficiently

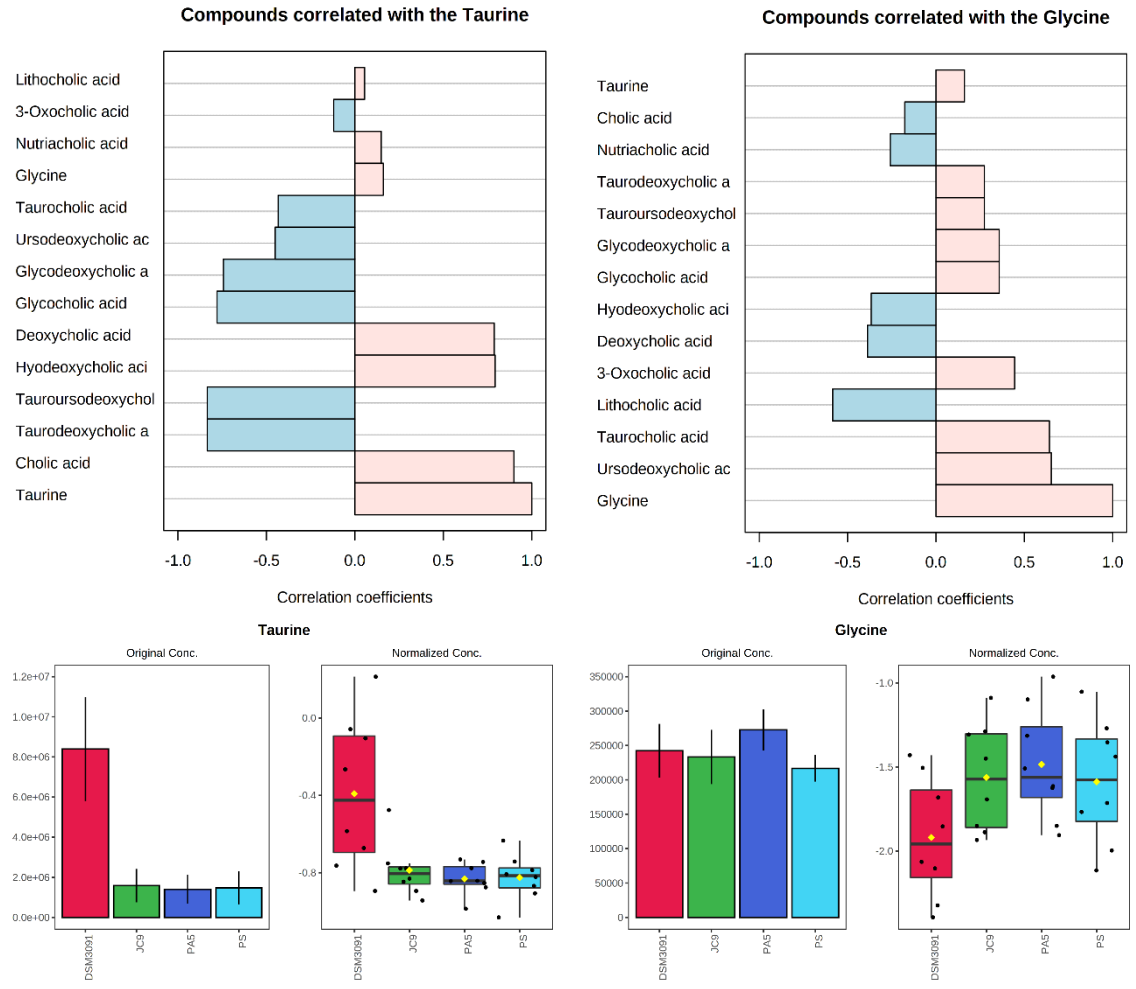


Figure 2.9. Correlations between bile acid metabolite profiles and the concentration of taurine and glycine. ANOVA was used to compare the taurine and glycine concentrations between each of the strains. Samples with 0% supplemented bile salts were removed from the analysis. Taurine was correlated with deconjugated bile acids, where glycine was correlated with conjugated bile acids. No statistically significant changes in taurine or glycine concentrations were observed between strains.

metabolised in culture with 0.1 and 0.25% bile salt supplementation. Taurocholic acid appeared to be efficiently metabolised in samples containing 1.0% supplemented bile acids. Collectively this suggests *M. stadtmanae* DSMZ3091 has differential efficiency in deconjugation for glyco- and tauro-

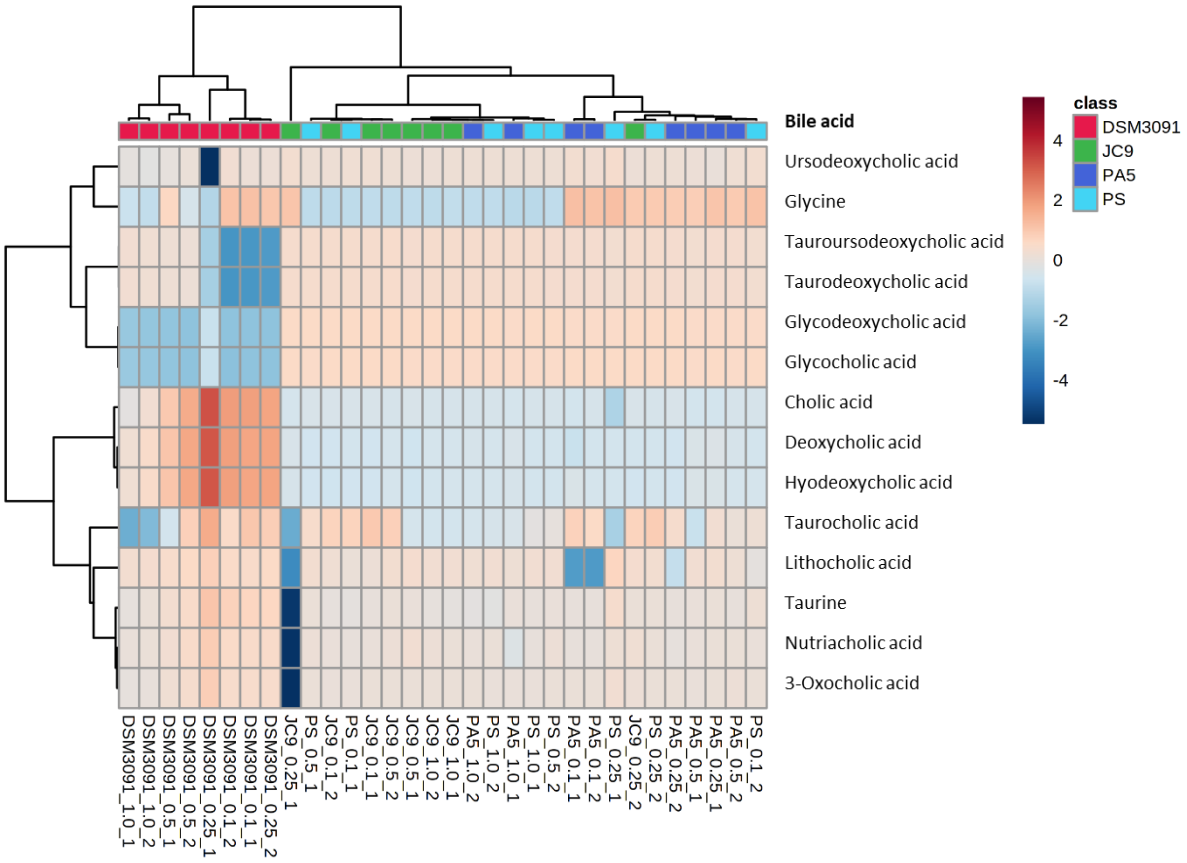


Figure 2.10. Heatmap displaying the hierarchical clustering of detected bile acid-associated metabolites in culture supernatants. Samples supplemented with 0% bile salt were removed from the analysis. Uninoculated medium metabolite concentrations were removed from respective culture supernatants. As shown in the heatmap, the bile acid metabolite profiles for *M. stadtmanae* DSMZ3091 are differentiated from the other strains.

conjugated bile acids at different concentrations.

2.3.5 The *bsh* of *M. smithii* and *M. stadtmanae* is constitutively expressed *in vitro*

With the differences observed in the bile acid metabolite profiles, I assessed the *bsh* gene expression of the *M. smithii* and *M. stadtmanae* supplemented with bile acids. As displayed in Figure 2.11, amplification of the *bsh* gene was observed at all bile salt concentrations for all strains of *M. smithii* and *M. stadtmanae*. This suggests that *bsh* is constitutively expressed for *M. smithii* and *M. stadtmanae in vitro*, even in cultures without additional bile salt supplementation. It is worth noting, as shown in Section 2.3.3, basal BRN-RF10 medium contained trace concentrations of taurocholic

acid, tauroursodeoxycholic acid, glycocholic acid, and taurodeoxycholic acid, in addition to a high concentration of ursodeoxycholic acid.

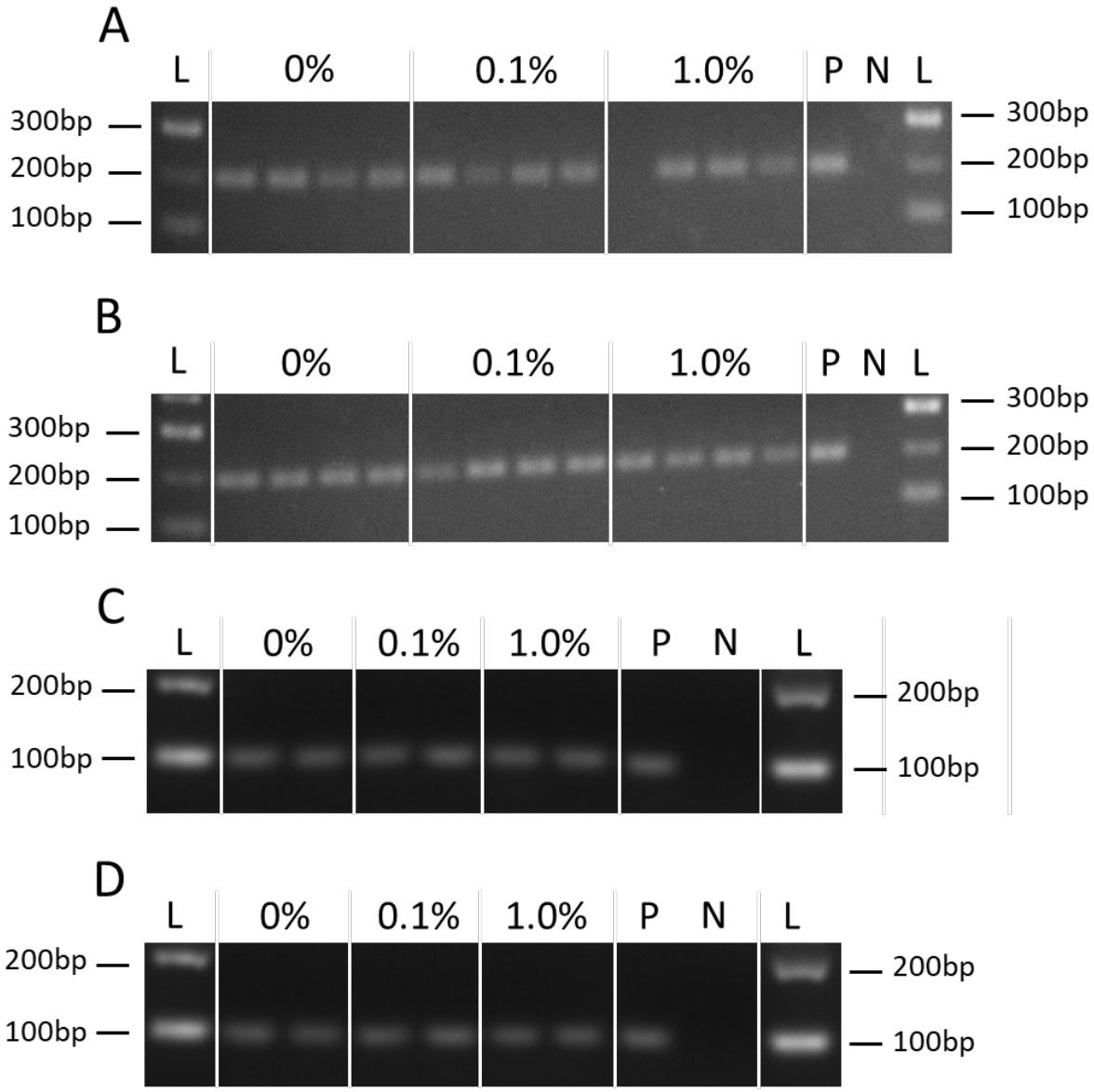


Figure 2.11. Gel electrophoresis showing the constitutive expression of *bsh* by *M. smithii* and *M. stadtmanae*. cDNA produced from cultures grown with supplemented bile salt (Figure 2.6) showed *bsh* expression at 0, 0.1 and 1.0%. **C)** and **D)** were run on the same gel and as such the ladder, positive and negative controls were the same. *Methanobrevibacter*- and *Methanosphaera*-specific *bsh* primers were designed to produce 83bp and 192bp amplicons, respectively. **A)** *M. stadtmanae* DSMZ3091, **B)** *M. stadtmanae* PA5, **C)** *M. smithii* PS, **D)** *M. smithii* JC9. P = positive control (respective genomic DNA), N = negative control (RNA sample without reverse transcription) and L = 100 bp Bioline HyperLadder.

2.3.6 *M. stadtmanae* PA5 shows significantly greater *bsh* expression compared to *M. stadtmanae* DSMZ3091

Annotation of the *M. stadtmanae* PA5 genome by Prokka produced two short sequences with predicted *bsh* activity. In order to confirm that the *bsh* of PA5 is transcribed into a complete mRNA sequence, primers were designed to 'bridge' the predicted region of truncation between the two *bsh* sequences (Figure 2.12A). These primers were used to query the cDNA produced from RNA extracted from the 1.0% supplemented cultures of *M. stadtmanae*, assuming *bsh* activity would likely be the highest in these samples. Amplification was achieved in samples produced from both *M. stadtmanae* DSMZ3091 and PA5 (Figure 2.12B, D). This likely suggests that the mRNA encoding for the *M. stadtmanae* PA5 *bsh* is produced as a single transcript despite the predicted truncation. To validate the sequence of the *M. stadtmanae* PA5 *bsh*, amplicons produced from cDNA were sequenced and compared to the *bsh* sequence from the respective genomes. Both the *M. stadtmanae* PA5 cDNA amplicon and genome sequence contained a single guanine nucleotide insertion at position 353 (of the PA5 *bsh*), validating the gene structure and confirming the presence of the

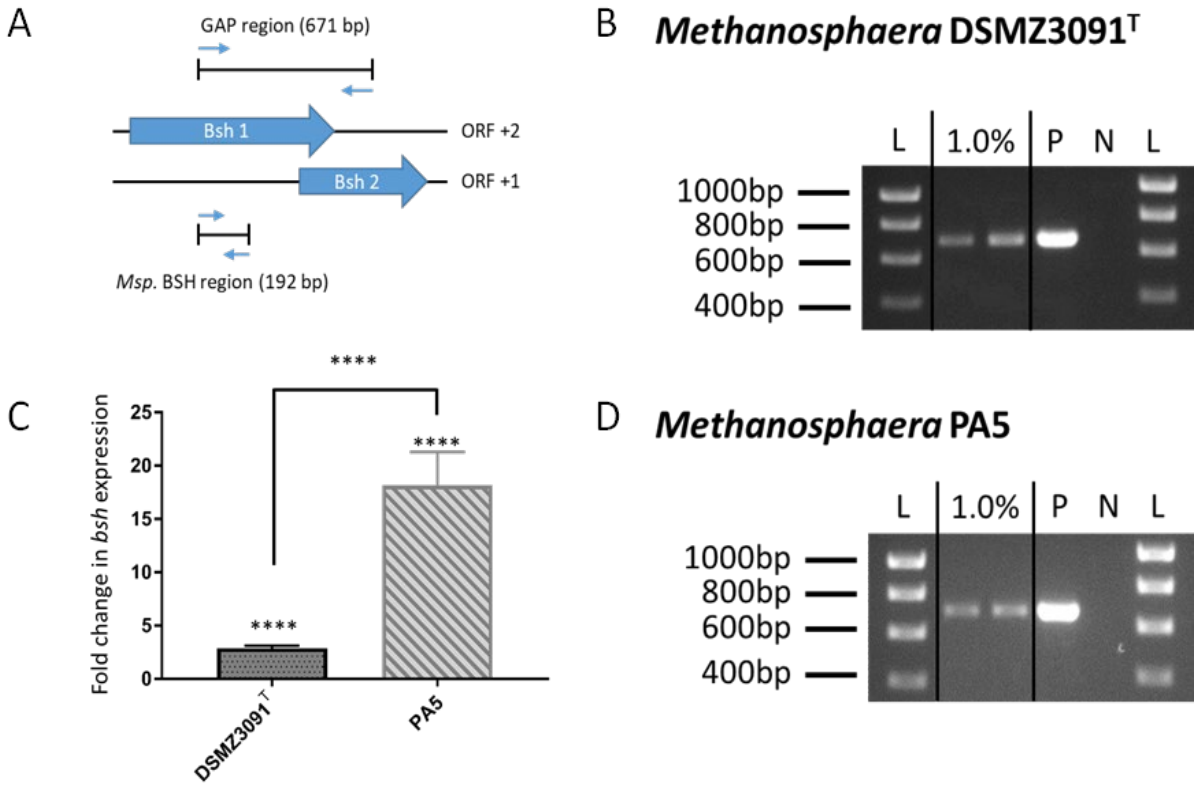


Figure 2.12. Structure and expression of the *bsh* of *M. stadtmanae* PA5 and *M. stadtmanae* DSMZ3091. A) Primers were designed to bridge the predicted region of *bsh* truncation for *M. stadtmanae* PA5. **B)** and **D)** show amplification of cDNA from *M. stadtmanae* DSMZ3091 and PA5 cultures supplemented with 1.0% bile acid. **C)** show fold change of *M. stadtmanae bsh* expression in 1.0% supplemented bile acid cultures relative to cultures with 0.0% using RT-qPCR.

insertion in the *bsh* mRNA (Figure 2.13). This would suggest some form of +1 ribosomal frameshifting is required to produce the complete *bsh* transcript. In addition to *M. stadtmanae* PA5, 24 MAGs contained *bsh* length discrepancies, seven with nucleotide insertions and 16 with nucleotide deletions were detected across various lineages of *Methanobrevibacter* and *Methanosphaera* (see Section 2.3.7). Like *M. stadtmanae* PA5, nucleotide insertions were shown in *bsh* of *M. smithii*, as well as *Methanobrevibacter* sp. YE315**.** *Bsh* sequences with nucleotide deletions were observed in *Methanobrevibacter wolinii,* human *Methanobrevibacter* spp., several *M. smithii,* and bovine-associated *Methanosphaera* lineages.

Figure 2.13. Partial *bsh* nucleotide sequence alignment of *M. stadtmanae* DSMZ3091 and PA5 *bsh* amplicons. *Bsh* nucleotides 318 to 361/362 are included for cDNA amplicons and their respective genome sequences. The single guanine nucleotide insertion is shown in the *M. stadtmanae* PA5 cDNA amplicon and genome sequence but not found in *M. stadtmanae* DSMZ3091.

Primers were designed to amplify the *bsh* of *M. stadtmanae* and used to quantify the fold change in gene expression with increasing bile acid concentrations (Figure 2.12A, Table 6.3). As shown in Figure 2.12C, both *M. stadtmanae* DSMZ3091 and PA5 showed a significant fold increase in *bsh* expression in cultures supplemented with bile salt compared to those without. However, *M. stadtmanae* PA5 showed a significantly greater increase at ~17-fold compared to the ~3-fold increase in *M. stadtmanae* DSMZ3091. This variation in *bsh* expression of *M. stadtmanae* highlights the intra-species variation in bile acid metabolism, though further work is necessary to confirm how this *bsh* gene structure and expression translate to *in vivo* bile acid metabolism within the gastrointestinal tract.

2.3.7 Phylogenetic assessment of the *bsh* of methanogenic archaea

Of the 572 compiled MAGs, 89% was found to possess at least one *bsh* homolog with >64% sequence identity to at least one of the *bsh* reference sequences. *Bsh* homologs were prevalent across *Methanobrevibacter*, *Methanosphaera,* and *Methanomassiliicoccales* lineages (Figure 2.14). Only one of the 451 genomes assigned to *M. smithii* lacked a *bsh* homolog, though the oral isolate *M. oralis* did not contain a *bsh* homolog. In contrast, *bsh* prevalence in the *Methanomassiliicoccales* and *Methanosphaera* lineages was more sporadic. Both the *M. luminyensis* and *M. intestinalis* genomes

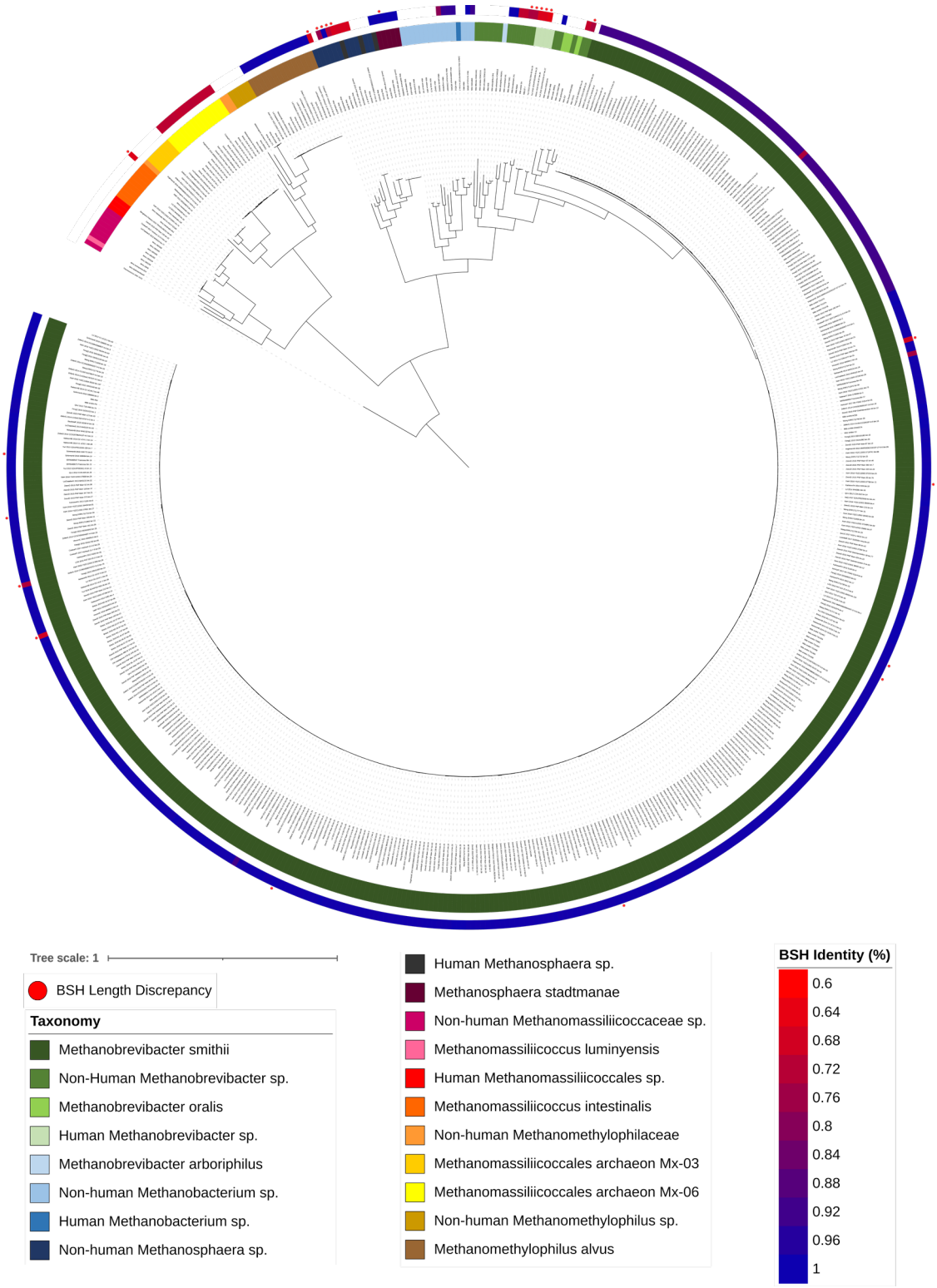


Figure 2.14. *bsh* homologs recovered from methanogen MAGs and isolate genomes. Taxonomic classification was performed using GTDB-tk (v1.3.0) and annotated as per the legend. *Bsh* homologs were identified using Kaptive (v0.5.1) with a custom database of methanogen *bsh* sequences. The percentage identity of *bsh* homologs to the database *bsh* is shown from 60-100%, as per legend. *Bsh* sequences with length discrepancies were annotated with a red circle (n=42). As the figure shows, *bsh* homologs are found in all major families of human-associated methanogens.

lacked a *bsh*, with only one of the *M. intestinalis* MAGs (HMP_2012__SRS048870__bin.16)

containing a homolog. Comparatively, all genomes assigned to *Methanomethylophilus alvus* and Methanomassiliicoccales archaeon Mx-06 (14/14) contained a *bsh* homolog, with only the Methanomassiliicoccales archaeon Mx-03 lineage being the exception. With reference to *Methanosphaera* genomes, *bsh* homologs were found within several human and bovine *Methanosphaera* genomes. However, some *Methanosphaera* from sheep, human, bovine and kangaroo did not contain *bsh*. Recently, Pasolli et al. (2019) produced human-associated MAGs associated with *Methanobacterium*, *Methanosphaera cuniculi*, non-human animal-associated *Methanobrevibacter*, and high-quality MAGs of the Methanomassiliicoccales archaeon Mx-06 lineage. Consequently, novel *bsh* sequences were also recovered from these genomes. It is worth noting the absence of *bsh* genes in species with limited representatives require validation by higher quality genomic assemblies.

Different types of length discrepancies can also be found in various *bsh* sequences, with the majority observed in species of non-human *Methanosphaera* and non-human *Methanobrevibacter* spp. Of the human strains, length discrepancies are found within 10 *M. smithii*, including two cultured isolates recovered as a part of a larger study on the pan-genome of *M. smithii* (Hansen et al., 2011). As previously stated, *M. stadtmanae* PA5 also appears to contain a length discrepancy (Hoedt et al., 2018). Given that *M. stadtmanae* PA5 represents only the second cultured isolate of human *M. stadtmanae*, further comparison between *M. stadtmanae* DSMZ3091^{T} and *M. stadtmanae* PA5 is of particular interest to determine the effect of this novel *bsh* gene structure.

Phylogenetic analysis of *bsh* showed stratification based on species and genus (Figure 2.15). This delineation was consistent between amino acid and nucleotide sequence phylogeny. Interestingly, there appeared to be multiple distinct clusters within the *bsh* of *M. smithii*. This separation is consistent with the phylogenetic differentiation observed between *M. smithii* and the largely uncharacterised subgroup *M. smithii*_A. Additionally, there were clear separations within each group, with *M. smithii* sequences separating into four phylogroups and *M. smithii*_A into three phylogroups (Figure 2.15). Although *M. arboriphilus* did contain a predicted *bsh* homolog according to the NCBI genome annotation, this gene clustered with predicted bacterial penicillin v-acylase sequences. Additionally, it was not recovered from the kaptive analysis and as such, it was excluded from further analyses.

Previously studies have suggested the *bsh* of methanogens was acquired through horizontal gene transfer from bacteria (Jones et al., 2008). To this end, phylogenetic analysis of the methanogen *bsh* shows the different lineages of *bsh* to cluster with those of different bacterial species (Figure 2.15). *Methanosphaera* and *Methanobrevibacte*r cluster closely with sequences from the order *Bacillales,* such as *L. monocytogenes, S. equinus, P. antarcticus,* and *E. faecium*. Comparatively, M. archaeon

Mx-06 and *M. alvus* cluster more closely with *L. delbrueckii*, *E. cellulosolvens,* and *S. variabile*. The *Methanobacterium* and *M. cuniculi bsh* more closely align with species of *Clostridium*, *C. butyricum,* and *C. lentocellum*, respectively. This together suggests these lineages have acquired their respective *bsh* through unique horizontal gene transfer events.

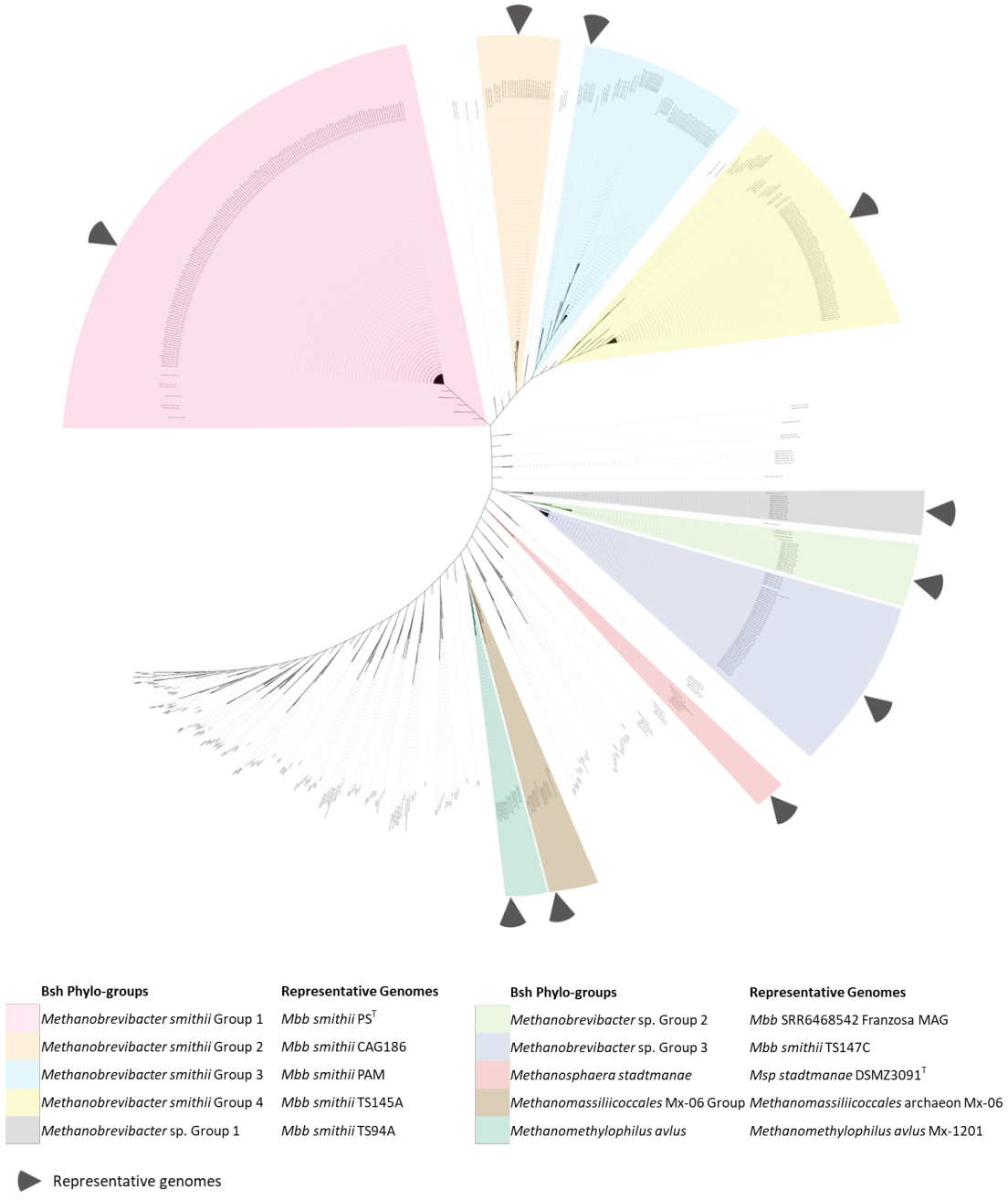


Figure 2.15. Phylogenetic analysis of predicted methanogen *bsh* homologs. *Bsh* were identified using Kaptive (v0.5.1) with a custom database of methanogen *bsh* sequences. *Bsh* phylogeny was analysed using ClustalO and MEGA-X, with Maximum likelihood and 1000 bootstraps. *Bsh* phylo-groups were coloured as per the legend. The position of each representative genome within their respective clade is denoted by the black arrow. Multiple phylogroups of *bsh* can be seen across the methanogen lineages.

One *M. smithii* (VincentC 2016 MM002.2 bin.14) and one *M. intestinalis* (HMP 2012 SRS048870 bin.16) MAG produced *bsh* homologs that were potentially erroneous. The former was truncated to 100 bps and the latter showed near-identical (99%) sequence similarity to *Anaerostipes hadrus*. Although this sequence may represent a recent gene acquisition, it more likely represents bacterial contamination of the MAG.

Conservation of the sequence was observed between the Bsh of all *M. smithii* clades, with regions of divergence observed between amino acids 110-134 and the C-terminus (Figure 2.16). These differences were similarly observed in the *bsh* nucleotide sequences (Figure 6.3). *M. stadtmanae, M.*

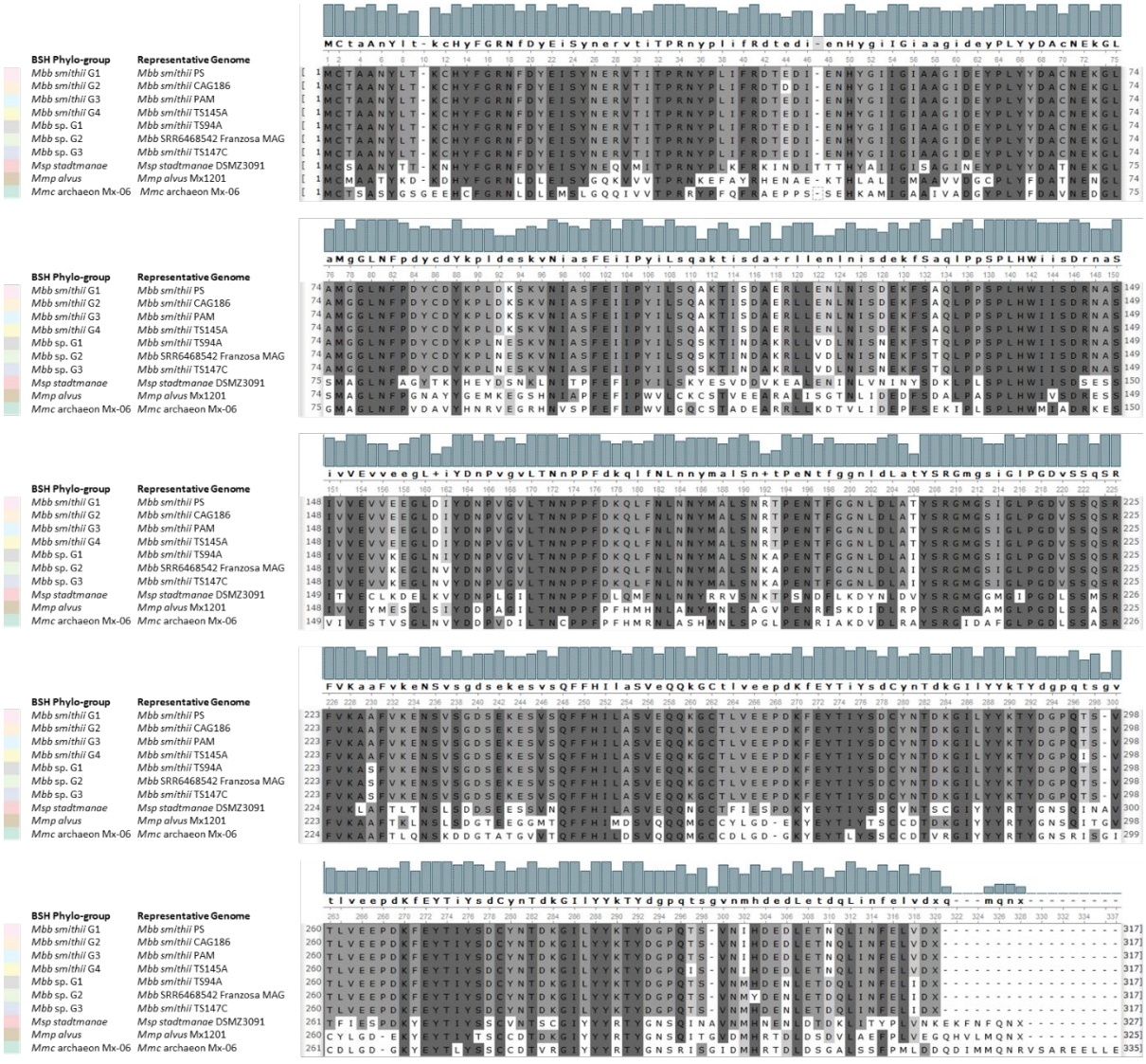


Figure 2.16. Multiple alignment of Bsh amino acids sequences from representative genomes of *bsh* phylogroups. *Bsh* phylogroups and representative genomes were determined as per Figure 2.15. The multiple alignment of Bsh was performed using MUSCLE and visualised with UGENE. Bsh sequences from *M. stadtmanae* and *Methanomassiliicoccales* lineages showed significant amino acid substitutions compared to *M. smithii* bsh.

alvus, and M. archaeon Mx-06 were divergent from the *M. smithii* sequences, with the latter two being more similar than *M. stadtmanae.* Analysis of the predicted catalytic amino acid residues showed Arg16, Asp19, Asn79, Asn170, and Arg223 to be conserved in *M. alvus* and *Methanobrevibacter* lineages, similar to those seen in *L. salivarius* strains (Table 2.2)(Fang et al., 2009). Conversely, M. archaeon Mx-06 and *Methanosphaera* show catalytic amino acid residues Arg16, Asp19, Asn80, Asn171, and Arg224, and Arg17, Asp20, Asn80, Asn171, and Arg224, respectively (Table 2.2). Residues Asn171 and Arg224 show conservation similar to strains of *E. faecalis* (Chand, Panigrahi et al. 2018). Catalytic activity and substrate specificity of Bsh have been linked to Asn79 (Öztürk & Önal, 2019) and all sequences contain a conserved Cys2, necessary for N-terminal nucleophile hydrolases (Lodola et al., 2012). These analyses show differential phylogeny and sequence conservation of the Bsh of different human methanogen lineages. This may translate to different enzymatic activity or substrate specificity to different bile acids, affecting their metabolism

Table 2.2. Conserved predicted catalytic residues of representative human methanogen Bsh protein sequences.

Phylogroup	Genome Representative	Predicted Catalytic Residues					
***Methanobrevibacter smithii* Group 1**	*M. smithii* DSMZ861 (PS)	Cys2	Arg16	Asp19	Asn79	Asn170	Arg223
***Methanobrevibacter smithii* Group 2**	*M. smithii* CAG186	Cys2	Arg16	Asp19	Asn79	Asn170	Arg223
***Methanobrevibacter smithii* Group 3**	*M. smithii* PAM	Cys2	Arg16	Asp19	Asn79	Asn170	Arg223
***Methanobrevibacter smithii* Group 4**	*M. smithii* TS145A	Cys2	Arg16	Asp19	Asn79	Asn170	Arg223
***Methanobrevibacter* sp. Group 1**	*M. smithii* TS94A	Cys2	Arg16	Asp19	Asn79	Asn170	Arg223
***Methanobrevibacter* sp. Group 2**	*M.* SRR6468542_Franzosa MAG	Cys2	Arg16	Asp19	Asn79	Asn170	Arg223
***Methanobrevibacter* sp. Group 3**	*M. smithii* TS147C	Cys2	Arg16	Asp19	Asn79	Asn170	Arg223
Methanosphaera stadtmanae	*M. stadtmanae* DSMZ3091 (MCB-3)	Cys2	Arg16	Asp19	Asn80	Asn171	Arg224
Methanomethylophilus alvus	*M. alvus* Mx05	Cys2	Arg16	Asp19	Asn79	Asn170	Arg223
Methanomassiliicoccales archaeon Mx-06	*M.* archaeon Mx-06	Cys2	Arg17	Asp20	Asn80	Asn171	Arg224

in the gastrointestinal tract. However, further characterisation is required to confirm these findings.

2.4 Discussion

M. smithii and *M. stadtmanae* are frequently described as members of the GIT microbiome, but little is known about their relationship with changing get ecology and predominant GIT metabolites. To my understanding, no other study has shown the effect of bile salts on the growth kinetics of human methanogens. *M. smithii* PS and *M. smithii* JC9 showed a similar tolerance to bile salts, with only higher concentrations (0.5-1%) showing an effect on maximum culture yield. Comparatively, substantial variation was observed between the tested *M. stadtmanae* strains. Indeed, *M. stadtmanae* PA5 showed no significant change in growth kinetics in response to bile salt supplementation, where *M. stadtmanae* DSMZ3091 showed a decrease in growth rate and an increase lag time with increasing bile concentrations. This work represents the first to show the differential *in vitro* bile salt tolerance of *M. smithii* and *M. stadtmanae*.

Previous analysis of the *M. smithii* PS *bsh* gene conducted by Jones et al. (2008) showed the recombinant gene to deconjugate both taurodeoxycholic and glycodeoxycholic acid when expressed in *E. coli*. My analysis of polar metabolites in culture supernatants showed an increase in cholic acid in all tested strains, suggesting *M. smithii* can also effectively deconjugate tauro- and glycocholic acids. Although the constitutive *bsh* expression was observed across *M. smithii* and *M.* stadtmanae, *M. stadtmanae* DSMZ3091 showed significantly higher potential for bile salt metabolism that has not previously been reported. Additionally, significant inter-genus and intra-species variation was observed under the tested culture conditions. Although the profiles of all detected metabolites clustered according to species, *M. stadtmanae* DSMZ3091 displayed a bile acid metabolite profile that was significantly different from all other tested strains. One of the metabolites responsible for this separation of *M. stadtmanae* DSMZ3091 was hyodeoxycholic acid, which was not previously associated with *M. stadtmanae* or *M. smithii*. This bile acid could be produced through the deconjugation of glyco- and taurohyodeoxycholic or the dehydroxylation of hyocholic acid, as show in the bacterium HDCA-1 observed in the intestine of a rat (Eyssen et al., 1999). Hyocholic acid and conjugated hyodeoxycholic acids were not observed by LC-QToF, thus further work is required to determine the pathway of hyodeoxycholic acid production by *M. stadtmanae* DSMZ3091. Predicted hydroxysteroid dehydrogenase (*hsdh*) genes, with predicted 7/12α/b-*hsdh* activity, have been recovered from *Methanobrevibacter* and *M. stadtmanae* based on homology to bacterial *hsdh*, suggesting these methanogens may contain a greater potential for bile acid metabolism (Doden et al., 2018; Heinken et al., 2019; Kisiela et al., 2012). Similarly, modules of predicted bile acid metabolism produced by Heinken et al. (2019) recovered 12-alpha/beta-hydroxysteroid dehydrogenase from *M. stadtmanae* and *M. smithii*. Although the methanogens were not predicted to produce secondary bile acids alone, both were predicted to produce ‘synthesis-enabled bile salt metabolites’, in which they

form a complementary pairing with a bacterium to produce secondary bile acids. Predicted pairs included *Mycobacterium, Plesiomonas*, *Rhodococcus,* and *Shigella*, and may provide *M. smithii* and *M. stadtmanae* with an increased bile acid metabolism potential in bacterial co-culture or as a part of the wider GIT microbiome. Along with the metabolomic analyses, I have shown greater potential capacity of *M. smithii* and *M. stadtmanae* to metabolise bile acids beyond simple deconjugation and the production of more diverse metabolites under a complex microbial environment.

To my understanding, these analyses similarly provided the first metabolomics analysis of *M. smithii* and *M. stadtmanae* grown *in vitro*. My analysis of bile acid metabolites by LC-QToF detected of 13 bile acids. This provided sufficient metabolites to investigate the deconjugation of conjugated cholic, deoxycholic and ursodeoxycholic acid. However, my analysis did not detect many bile acid derivates, including chenodeoxycholic acid and the associated conjugates. In fact, these bile acids accounts for ~40% of the human bile acid pool, equal to that of cholic acid (~40%) and more than deoxycholic acid at ~20% (Chiang, 2017). Previous analysis of Oxoid bile salt preparations showed chenodeoxycholic acid and its conjugates to be present, albeit at lower concentrations compared to cholic and deoxycholic acids (Hu et al., 2018). As such, further work should focus on the effect of other important human-associated bile acids, such as chenodeoxycholic acid, on the growth and metabolism of *M. smithii* and *M. stadtmanae*.

In addition to the intra-*M. stadtmanae* variation in bile acid metabolite profiles, *M. stadtmanae* PA5 contained a unique *bsh* gene structure compared to *M. stadtmanae* DSMZ3091. The annotation of the genome predicted a truncation of the *bsh* gene resulting from a single guanine insertion. This mutation was confirmed through sequencing of cDNA transcripts produced from the total RNA of *M. stadtmanae* PA5 cultures grown with 1.0% bile salt supplementation. Additionally, specific primers designed to 'bridge' the predicted region of truncation produced single transcripts in both *M. stadtmanae* PA5 and *M. stadtmanae* DSMZ3091. This suggests a unique mechanism of transcription or translation is required for *M. stadtmanae* PA5 to produce a functional Bsh protein. Translational read-throughs and frame-shifting are well documented in methanogens, such as in the methylamine-utilising *Methanomassiliicoccus* (Borrel et al., 2014). The methylamine transferase genes of these species encode an in-frame amber stop codon which encodes for the insertion of pyrrolysine and produces a single mRNA transcript, where termination of the transcription would otherwise occur (Antonov et al., 2013). Additionally, the synthesis of magnesium chelatase from *Methanocaldococcus* and *Methanococcus* is also predicted to require -1 translational frame-shift (Antonov et al., 2013). Comparatively, the *bsh* of *M. stadtmanae* PA5 would require a predicted +1 frameshift to produce a functional mRNA transcript. This type of frame shift is scarcely explored in archaeal species, though it has been observed in myovirus of the halophilic archaea *Halorubrum*

sodomense and *Haloarcula sinaiiensis* (Pietilä et al., 2013; Senčilo et al., 2013). Interestingly, my analyses showed additional species of *Methanobrevibacter* and *Methanosphaera* also contained *bsh* length discrepancies, such as the bovine isolate *Methanosphaera* sp. BMS, suggesting that mutation of the *bsh* gene may be more common than expected across methanogen species.

Interestingly, the *bsh* expression of *M. stadtmanae* DSMZ3091 and *M. stadtmanae* PA5 showed the latter to produce significantly high fold-change in expression in cultures supplemented with bile salt. Further work is required to characterise this potentially novel translational frame-shifting and the effect on the expression of the *bsh* of *M. stadtmanae* PA5 and other methanogen species. As shown in Section 2.3.3, the basal BRN-RF10 medium contained trace concentrations of taurocholic acid, tauroursodeoxycholic acid, 3-oxocholic acid, glycocholic acid, and taurodeoxycholic acid, as well as a high concentration of ursodeoxycholic acid. The constitutive expression of *bsh* in cultures of *M. smithii* and *M. stadtmanae* may be attributed to this trace concentration of bile acid derivatives. Future studies should focus on the use of a medium devoid of bile acids, likely achieved through the removal of rumen fluid from the basal medium. This would provide a more accurate comparison of bile acid expression between cultures with and without bile acid supplementation.

In conclusion, *M. smithii* and *M. stadtmanae* show a significant diversity of bile salt metabolism and tolerance. *M. stadtmanae* PA5 and *M. smithii* were largely unaffected by bile acid supplementation, with the *M. smithii* strains only showing deceased yield at higher bile salt concentrations. Comparatively, *M. stadtmanae* DSMZ3091 showed a decreased growth rate with increased bile acid concentration. Analysis of polar metabolites within the culture supernatants showed *M. stadtmanae* DSMZ3091 to produce a significantly different bile acid metabolite profile compared to the other strains. Additionally, *M. stadtmanae* PA5 contained a novel *bsh* gene structure and gene expression compared with *M. smithii* and *M. stadtmanae* DSMZ3091. These results collectively show a diversity of bile salt tolerance and metabolism in *M. smithii* and *M. stadtmanae* strains, which may play a substantial role in the response to the altered bile acid pool and gut ecology observed in gastrointestinal disease.

2.5 Published research articles on work carried out in Chapter 2

Enrichment, isolation, and preliminary characterisation of *M. stadtmanae* PA5 was published as a part of a larger study on the culture and metagenomic analyses of *Methanosphaera*. Additional work was also conducted with *M. stadtmanae* PA5 and included in the paper that was not described in Chapter two.

The ISME Journal (2018) 12:2942–2953
https://doi.org/10.1038/s41396-018-0225-7

ISME

ARTICLE

Culture- and metagenomics-enabled analyses of the *Methanosphaera* genus reveals their monophyletic origin and differentiation according to genome size

Emily C. Hoedt[1] · Donovan H. Parks[2,3] · James G. Volmer[1,3] · Carly P. Rosewarne[4] · Stuart E. Denman[5] · Christopher S. McSweeney[5] · Jane G. Muir[6] · Peter R. Gibson[6] · Páraic Ó Cuív[1] · Philip Hugenholtz[1,2,3] · Gene W. Tyson[2,3] · Mark Morrison[1,2]

Received: 27 November 2017 / Revised: 27 April 2018 / Accepted: 3 June 2018 / Published online: 1 August 2018

Abstract

The genus *Methanosphaera* is a well-recognized but poorly characterized member of the mammalian gut microbiome, and distinctive from *Methanobrevibacter smithii* for its ability to induce a pro-inflammatory response in humans. Here we have used a combination of culture- and metagenomics-based approaches to expand the representation and information for the genus, which has supported the examination of their phylogeny and physiological capacity. Novel isolates of the genus *Methanosphaera* were recovered from bovine rumen digesta and human stool, with the bovine isolate remarkable for its large genome size relative to other *Methanosphaera* isolates from monogastric hosts. To substantiate this observation, we then recovered seven high-quality *Methanosphaera*-affiliated population genomes from ruminant and human gut metagenomic datasets. Our analyses confirm a monophyletic origin of *Methanosphaera* spp. and that the colonization of monogastric and ruminant hosts favors representatives of the genus with different genome sizes, reflecting differences in the genome content needed to persist in these different habitats.

An additional paper entitled ‘Genomic- and culture-based analyses of the bile salt metabolism of human methanogenic archaea’ has been drafted, with intended submission to ISME Journal in Q4 2021.

Chapter 3: Identification of novel phylogroups of human methanogens and an assessment of their genomic potential

3.1 Introduction

Methanogenic archaea are important members of the GIT microbiome and help maintain the efficiency of bacterial fermentation. Though several species of methanogen have been associated with the GIT, the most abundant and prevalent species of human-associated methanogen is *M. smithii*. In healthy individuals, *M. smithii* can account for up to 10% of the total GIT microbiome and has been detected in up to 96% of healthy individuals (Dridi et al., 2009; Dridi, Raoult, et al., 2011; Eckburg et al., 2005). Typically regarded as the second most dominant methanogen, *M. stadtmanae* is also present at a lower abundance and is detected in 30% of subjects (Dridi, Henry, et al., 2012). Comparatively, populations of *Methanomassiliicoccales* are also detected in 4%-50% of individuals, with a significant increase in abundance observed in elderly subjects (Dridi, Henry, et al., 2012; Vanderhaeghen et al., 2015). The sole documented energy production pathway for *M. smithii* is hydrogenotrophic methanogenesis, primarily involving the utilisation of carbon dioxide and hydrogen to produce methane and thereby support ATP generation and growth. This hydrogenotrophic methanogenesis is enabled by hydrogen (and carbon dioxide) resulting from bacterial fermentation and this "interspecies hydrogen transfer" maintains favourable conditions for bacterial fermentation that would otherwise be inhibited by the build-up of hydrogen (Nakamura et al., 2010).

Methanogen community profiles and abundance are variable in both healthy and diseased individuals. One study by Blais Lecours et al. (2014) showed a reduction in the typically dominant *M. smithii* and an increased prevalence of *M. stadtmanae* in patients with IBD. Notably, this study showed a significant *M. stadtmanae*-specific IgG immune response in patients with IBD compared to healthy subjects, with peripheral blood mononuclear cells (PBMCs) also producing a significant proinflammatory TNFα cytokine response when exposed to *M. stadtmanae* but not *M. smithii*. Similarly, a reduction in *M. smithii* was also observed in patients with IBD from British and Iranian studies (Ghavami et al., 2018; Scanlan et al., 2008). Additionally, a cohort of Russian subjects with IBD showed a lower abundance of *Euryarchaeota*, attributed to *M. smithii*, in patients with CD compared with UC, although patients with UC also showed a reduction of *Methanobrevibacter* compared to healthy subjects (Lo Sasso et al., 2020).

Similar reductions in *M. smithii* were observed in obese individuals, though an increase was observed in unclassified species of *Methanobrevibacter* (Maya-Lucas et al., 2018; Million et al., 2013; Million et al., 2012). Conversely, *Methanobrevibacter* spp. showed an increased abundance in subjects with

metabolic syndrome and anorexia (Armougom et al., 2009; Lim et al., 2017). Taken together, these studies show increasing evidence that changes in GIT ecology caused by disease can differentially affect the prevalence and abundance of methanogen species, although scant analyses have been conducted on the genomic potential of the different methanogen species. Samuel et al. (2007) provided an analysis of the genomic and metabolic adaptations of four available *M. smithii* strains. Although a limited number of strains were assessed, key differences were observed in the cell surface structure variation, as well as the potential formate utilisation for methanogenesis. Hansen et al. (2011) furthered these finding by the isolation and comparative analysis of 20 *M. smithii* strains from monozygotic and dizygotic twin pairs. 101 adhesin-like proteins (ALPs) were differentially encoded by the *M. smithii* strains and showed strain-specific differentiation in expression. As such, Hansen et al. (2011) hypothesise that these strain-specific ALP repertoires allow for syntrophic interactions with differential bacterial species.

During my research for this Chapter, Pasolli et al. (2019) reported their recovery of more than 150,000 MAGs of human origin from 9,428 metagenomes. Of the reported MAGs, 676 (0.45%) were assigned to methanogenic archaeal lineages, with ~90% of these assigned to the genus *Methanobrevibacter* and the remaining MAGs assigned to *Methanosphaera*, *Methanobacterium*, *Methanomassiliicoccus*, and *Methanoculleus*. Interestingly, 47 of the methanogen MAGs (6.9%) were also assigned to uncharacterised lineages of *Methanomassiliicoccaceae* and *Methanobacteriaceae*. Based on the availability of this expanded dataset, I have examined these human-derived MAGs along with MAGs recovered as a part of my analysis, and reference genomes available from the NCBI genome database. Together these MAGs and isolate genomes provide genomic information on diverse human methanogen species representing various geographical locations and health conditions. Using these MAGs, I have identified methanogen phylogroups associated with key GIT disease states and the enrichment of specific genes within these phylogroups of interest.

3.2 Materials and Methods

3.2.1 Recovery of methanogen MAGs from publicly available metagenome datasets

Publicly available human faecal metagenome datasets were downloaded from the NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra) for the attempted recovery of methanogen MAGs. Details of the studies analysed are included in Table 3.1.

Table 3.1. Reference information for metagenome datasets used to recover human methanogen MAGs as of 18/05/2019.

Author	BioProject	PMID	Reference
Franzosa_2019	PRJNA400072	30531976	Franzosa et al. (2019)
Wang_2018	ERP010708	30208875	Wang et al. (2018)
YeZ_2018	PRJNA356225	30077182	Ye et al. (2018)
WengY_2019	PRJNA429990	31240835	Weng et al. (2019)
GuitorA_2019	PRJNA540073	31611361	Guitor et al. (2019)
PhilipsA_2017	PRJNA354503	28609785	Philips et al. (2017)
HallA_2017	PRJNA385949	29183332	Hall et al. (2017)

Metagenome sequences were first trimmed using Trimmomatic (v0.32)(Bolger et al., 2014). MegatHit (v1.1.1)(Li et al., 2015) was used for the subsequent assembly and reads were mapped to the assembly using BamM (v1.7.3)(https://github.com/Ecogenomics/BamM). Sequences were then binned using the 'superspecific' parameter of Metabat (v0.32.4)(Kang et al., 2015) and quality assessed using CheckM (v1.0.7)(Parks et al., 2015). Those which were classified as archaeal and had a completeness of ≥50% and contamination of ≤10% were retained. MAGs with completeness of ≥50 % and contamination of ≤10% were termed medium-quality (MQ). MAGs of ≥90% and ≤5% contamination were termed high-quality (HQ) and used for comparative analyses.

3.2.2 Quality assessment and phylogenetic analysis of methanogen MAGs and isolate genomes

Recovered methanogen MAGs were combined with those recovered and published by Pasolli et al. (2019). Genome statistics and quality scores were determined using CheckM (v1.0.7)(Parks et al., 2015). Concatenated archaeal marker gene files were produced and taxonomy was inferred using GTDB-tk (v1.3.0). Taxonomic classifications were visualised as Sankey diagrams using SankeyMATIC (www.sankeymatic.com/build/). Phylogenetic analysis of concatenated marker gene files was conducted using FastTree (v2.1.10) and visualised using iToL (www.itol.embl.de/). Geography and disease status was determined based on the respective metadata that was available, as

such metagenomes without necessary metadata were term 'N/S'. Geography and taxonomy were displayed as heatmaps using GraphPad Prism 9.

3.2.3 Comparative genomic analysis of *M. smithii* phylogroups using EnrichM

Comparative genomic analysis of methanogen MAGs and isolate genomes was conducted using EnrichM (v0.4.9) (https://github.com/geronimp/enrichM). The enrichM 'annotate' function was used to annotate KEGG Orthology (KO) using the '--ko' parameter and carbohydrate-active enzymes (Cazymes) using the '--cazy' parameter. The enrichment of genes in genome groups of interest was determined using the EnrichM 'enrichment' function with KO and Cazyme annotations. Statistical analysis of enriched genes was performed through EnrichM with Fisher's Exact test and Mann–Whitney U test. Values were visualised as heat maps on phylogenetic trees constructed as described in 3.2.2.

3.2.4 Recovery and phylogenetic analysis of *walc*- and *wald*-associated gene homologs using Kaptive

Kaptive (v0.5.1) (Wyres et al., 2016) was used to recover alcohol and aldehyde dehydrogenase gene homologs, associated with the utilisation of ethanol in *Methanosphaera* sp. WGK6 (Hoedt et al., 2016), from the human methanogen MAGs and isolate genomes. Two separate custom databases containing the '*walc*' (NL43_RS02830; WP_069592539.1) and '*wald*' (NL43_RS02835; WP_198923183.1) genes were used. Recovered sequences were translated into amino acid sequences using EMBOSS Transeq (Madeira et al., 2019; Rice et al., 2000). Phylogenetic analysis was performed using MEGA-X (Kumar et al., 2018). Sequences were aligned using MUSCLE, and phylogeny was inferred using Maximum-likelihood with the JJT model and 1000 bootstrap replications.

3.3 Results

3.3.1 Phylogenetic analysis of methanogen MAGs recovered from publicly available human faecal metagenome datasets

Seventy four MAGs of atleast medium-quality (MQ) were successfully recovered, with 75% (55/74) of these determined to be high-quality (HQ). As shown in Table 3.2, the vast majority of these MAGs (72/74) were derived from two datasets produced by Franzosa et al. (2019) and Wang et al. (2018). Over 90% of the MAGs (50/55) in both categories were taxonomically affiliated with the species *M. smithii.* Of the remaining MAGs, two HQ MAGs were assigned to *M. alvus* and the unclassified genus 'UBA71'; and two MQ MAGs were assigned to *M. intestinalis.* Interestingly, no MAGs recovered from these datasets were affiliated with the genus *Methanosphaera*. The HQ MAGs I produced were then combined with 467 HQ archaeal MAG produced by Pasolli et al. (2019). I also included 97 reference genomes produced from methanogen isolates of human and non-human origin (see Table 6.5 for a complete list of reference genomes). The reference genomes of human methanogens included 34 *Methanobrevibacter* (*smithii*, *oralis*, and *arboriphilus*), three *M. stadtmanae* (DSMZ3091^T, PA5, and the MAG DEW79), and four *Methanomassiliicoccales* genomes from *M. luminyensis*, *Ca.* Methanomassiliicoccus intestinalis, and *Ca.* Methanomethylophilus alvus (strains Mx05 and Mx1201^T). For additional reference, 28 genomes of methanogens from non-human animals were also included for *Methanobrevibacter*, *Methanosphaera, Methanobacterium, Methanomassiliicoccus, Ca.* Methanomethylophilus, and *Methanocorpusculum*. Given the identification of *Methanobacterium* and *Methanoculleus* MAGs produced by Pasolli et al. (2019), 12 *Methanobacterium* and 11 *Methanoculleus* genomes from environmental sources were also included as reference genomes in the analysis.

Table 3.2. Publicly available human MGS datasets used for the recovery of archaeal MAGs. Metagenome datasets were downloaded from the NCBI SRA database and archaeal MAGs were recovered as per the methods.

Author	BioProject	PMID	Samples Analysed	MAG Prefix ID	No. MQ MAGs	No. HQ MAGs
Franzosa_2019	PRJNA400072	30531976	270	Franzosa	6	14
Wang_2018	ERP010708	30208875	201	Wang	12	40
YeZ_2018	PRJNA356225	30077182	150	YeZ	0	1
WengY_2019	PRJNA429990	31240835	40	-	0	0
GuitorA_2019	PRJNA540073	31611361	6	GuitorA	1	0
PhilipsA_2017	PRJNA354503	28609785	165	-	0	0
HallA_2017	PRJNA385949	29183332	262	-	0	0

The vast majority of methanogen MAGs (424/472) were assigned to the genus *Methanobrevibacter,* with more than 80% assigned to the species *M. smithii.* Phylogenetic analysis of the combined *M. smithii* MAGs classified 19% (84/450) of *M. smithii* genomes to *M. smithii*_A, a distinct phylogroup defined by GTDB-tk. Interestingly, despite over 80% of the *M. smithii* genomes being assigned to the phylogroup *M. smithii*, there was an even number of cultured isolates from *M. smithii* and *M. smithii*_A lineages (Figure 3.1). Most of these genomes (labelled as 'TS') for both phylogroups can be attributed to a single study by Hansen et al. (2011). Interestingly, four unclassified *Methanobrevibacter* MAGs were recovered, one from a Fijian population and three from a Chinese population. Three of these MAGs did not cluster closely with any reference genomes, however, one

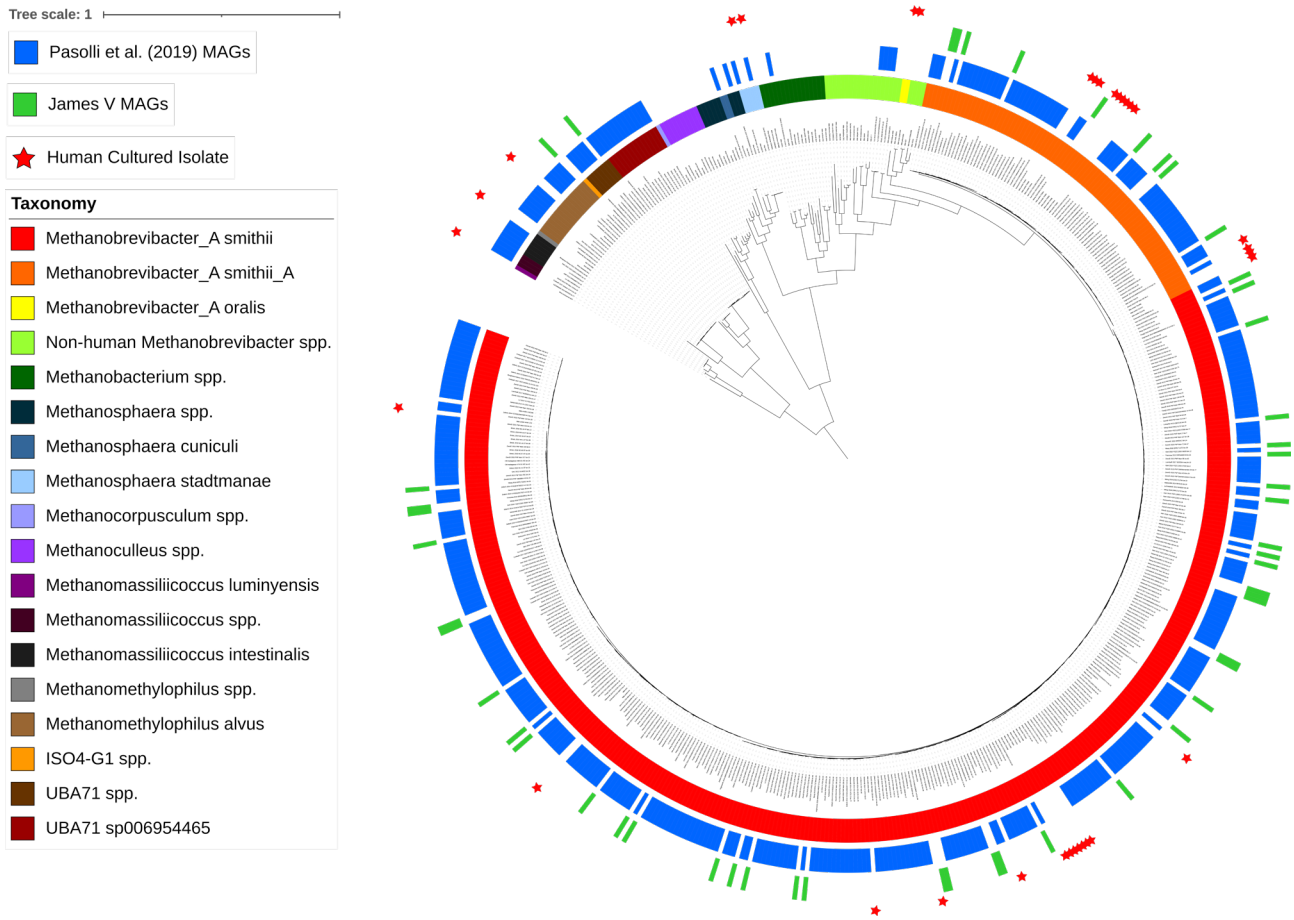


Figure 3.1. Phylogenetic distribution of human-associated methanogen MAGs and isolate genomes. HQ MAGs produced by Pasolli et al. (2019) were combined with the MAGs produced here. The phylogeny of human-derived methanogen MAGs and reference genomes was analysed using GTDB-tk (v1.3.0), FastTree (v2.1.10), and visualised using iToL (https://itol.embl.de/). The *Methanomassiliicoccales*-associated clade was chosen as the outgroup. MAGs recovered by Pasolli et al. (2019) and this analysis are displayed by the blue and green heatmaps, respectively. Genomes not recovered by Pasolli et al. (2019) or this analysis represent reference genomes. Human-derived cultured isolates are displayed by the red stars.

MAG was shown to be a human-derived representative of *Methanobrevibacter woesei*. As these MAGs were derived from MGS of human faecal samples, it was expected that no MAGs were assigned to the *Methanobrevibacter oralis*, which is typically the dominant oral *Methanobrevibacter* species.

Although thought to be the second most dominant methanogen, only three HQ MAGs from the combined dataset were classified to the genus *Methanosphaera*, with two of these MAGs assigned to the species *M. stadtmanae* and the other phylogenetically associated with *Methanosphaera* sp. WGK6, a *Methanosphaera* isolated from a Western grey kangaroo (Hoedt et al., 2016). Additionally, a MAG was classified by Pasolli et al. (2019) as *Methanobacteriaceae* at the family level, however my analysis showed it was phylogenetic similarity to the rabbit-associated lineage *Methanosphaera cuniculi*.

Interestingly, all predicted '*Methanoculleus*' MAGs produced by Pasolli et al. (2019) clustered with *Ca.* Methanomethylophilus alvus and not with any *Methanoculleus* reference genomes, suggesting that they were initially misclassified (Figure 3.1). For *Methanomassiliicoccus,* eight genomes were assigned to *Ca.* Methanomassiliicoccus intestinalis, with no recovered MAG was assigned to *Methanomassiliicoccus luminyensis* despite the fact that this species was previously recovered from human faeces(Dridi, Fardeau, et al., 2012). Twenty genomes, representing almost half of the recovered *Methanomassiliicoccales*-associated MAGs, contained no cultured representative and cluster closer with the ovine isolate Methanogen archaeon ISO4-G1 (Kelly et al., 2016). The largest cluster of these genomes (70%) were classified as Methanogenic archaeon Mx-06, which currently does not contain a cultured representative.

Given the overwhelming prevalence of MAGs and genomes representing *Methanobrevibacter* spp. available from these datasets, I will emphasise my findings from comparative genomics I have performed with the members of this lineage and include my findings with the non-*Methanobrevibacter* lineages where appropriate.

3.3.2 Geographical distribution of human-associated methanogen MAGs and isolate genomes

The combined HQ MAGs recovered from publicly available human metagenomes expand the geographical distribution of available human methanogen genomes to include 22 countries from six continents. As shown in Figure 3.2, the methanogens lineages were spread across various geographical locations, though there was a clustering of geographical locations within certain lineages.

Recovered MAGs and isolate genomes were dominated by Great Britain (21%) and Israel (19%), in addition to the USA (11%) and France (8%)(Figure 3.3). As expected, *M. smithii* was the most dominant methanogen for the vast majority (20/22) of geographical locations. Among the geographical location with a larger number of genomes, phylogroup *M. smithii*_A accounted for anywhere between 0 and 50% of the total *M. smithii* genomes, such as 47% of the USA genomes (Figure 3.3). Interestingly though, MAGs recovered from Spanish individuals were only classified to

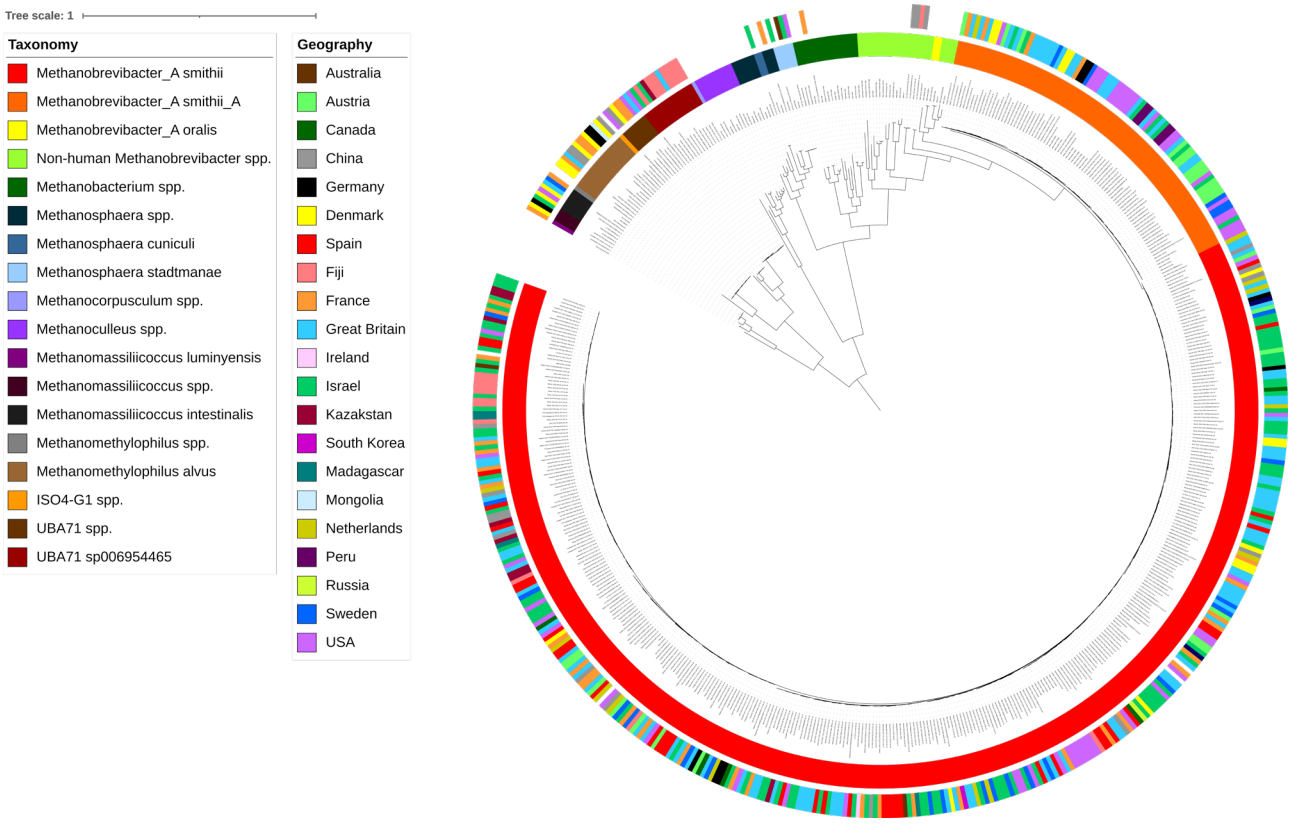


Figure 3.2. Geographical distribution of human-associated methanogen MAGs and isolate genomes. Phylogeny of human-derived methanogen MAGs and reference genomes was analysed using GTDB-tk (v1.3.0), FastTree (v2.1.10), and visualised using iToL (https://itol.embl.de/). The *Methanomassiliicoccales*-associated clade was chosen as the outgroup. Geographical locations of MAGs and isolate genomes was assigned according to available metadata and is indicated by the outer ring. The taxonomy of each MAG is displayed by the inner ring.

the phylogroup *M. smithii* and no other lineages. Comparatively, Austrian MAGs were evenly distributed between *M. smithii* and *M. smithii*_A, and MAGs recovered from Peru were only assigned to *M. smithii*_A. Although *M. smithii*_A is typically less dominant, there is a differential distribution of the *M. smithii* phylogroups across the geographical locations.

French metagenomes produced the highest diversity of taxonomic classifications with nine unique phylogenies and recovered the widest variety of MAGs assigned to lineages other than *M. smithii* or

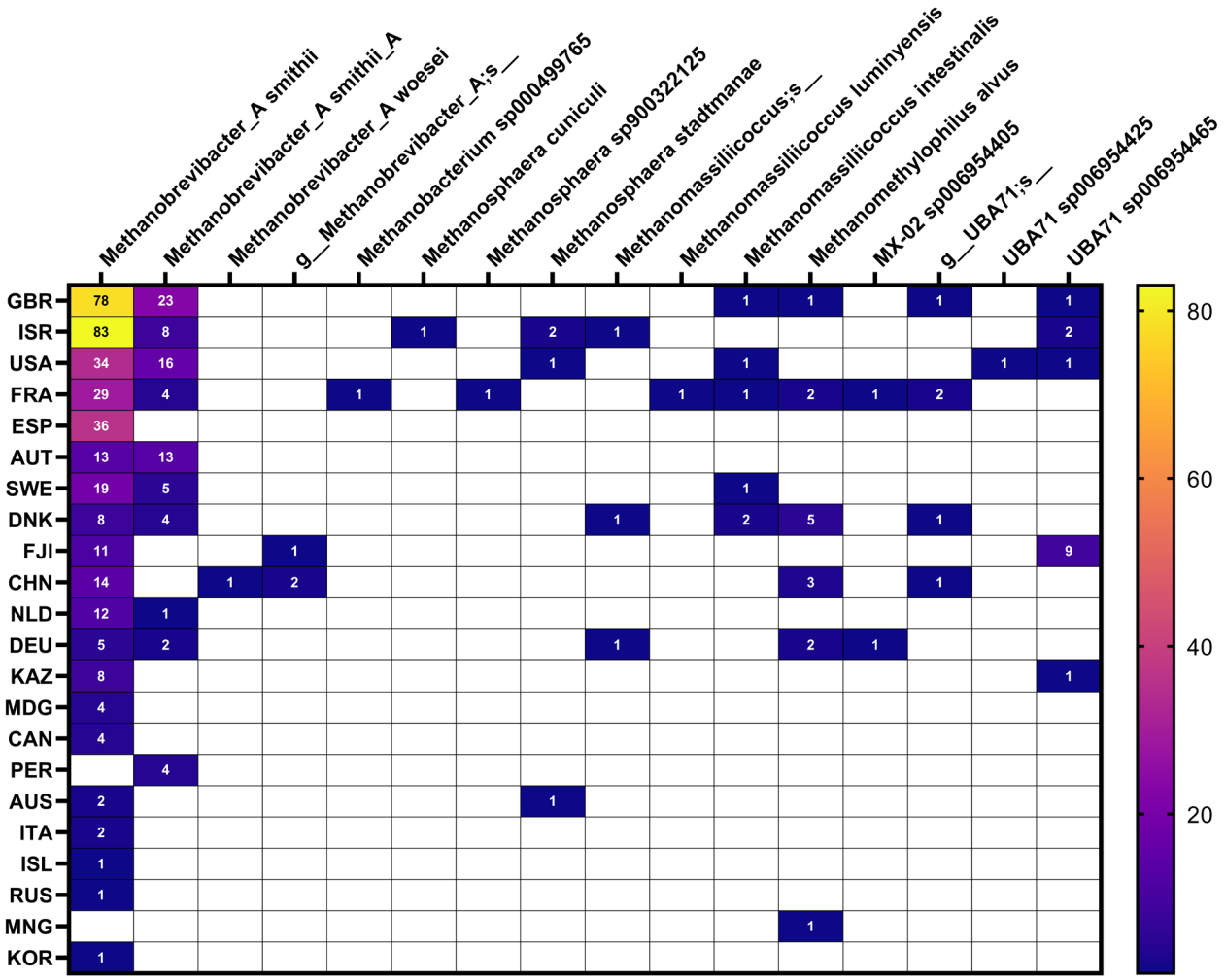


Figure 3.3. Heatmap displaying the geographical distribution of human-associated methanogen taxonomies. HQ MAGs were combined with human-derived cultured isolates for a total of 503 genomes with available geographical metadata from 22 geographical locations. Taxonomic classification was conducted using GTDB-tk (v1.3.0) and visualised using GraphPad Prism 9. *Ca.* M. alvus includes the reclassified genomes identified by Pasolli et al. (2019) as *Methanoculleus*. GBR, Great Britain; ISR, Israel; USA, United Stated of America; FRA, France; ESP, Spain; AUT, Austria; SWE, Sweden; DNA, Denmark; FJI, Fiji; CHN, China; NLD, Netherlands; DEU, Germany; KAZ, Kazakhstan, MDG, Madagascar; CAN, Canada; PER, Peru; AUS, Australia; ITA, Italy; ISL, Iceland; RUS, Russia; MNG, Mongolia; KOR, South Korea.

*M. smithii*_A (Figure 3.4). Interestingly, the MAGs assigned to the non-human *Methanobacterium* sp. MB1 (GTDB-tk classification sp000499765) and *Methanosphaera* sp. RUG761 (GTDB-tk classification sp900322125) lineages were also recovered from French samples. As mentioned in Section 3.3.1, representatives of unclassified *Methanobrevibacter*_A and *Methanobrevibacter*_A woesei were only recovered from Chinese and Fijian populations. Similarly, nearly 64% of MAGs assigned to UBA71 sp006954465 were recovered from Fijian samples, suggesting that lifestyle factors associated with these geographical populations, or geographical isolation in regards to Fiji, may promote the growth of different methanogen species other than the typical *M. smithii*. The only representative of *M. luminyensis* was that previously isolated by Dridi, Fardeau, et al. (2012). No representative MAG was recovered from any metagenome including the French samples, despite this species being previously cultured from the faeces of a French individual.

Interestingly, most of the available *M. smithii* isolate genomes were collected as a part of a study on the pan-genome of *M. smithii* in an American cohort of twins (Hansen et al., 2011). Althought the *M. smithii*_A phylogroup was not described in this study, 52% (14/27) of cultured *M. smithii* isolates were classified as *M. smithii*_A. In fact, 89% of these cultured isolates were recovered from the USA, with two Australian isolates recovered as a part of my analyses and *M. smithii* KB11 recovered from a South Korean individual (Kim & Jeong, 2018). Like *M. smithii*, the cultured representatives of *Methanomassiliicoccales* were also dominated by a single geographical location. In fact, cultured representatives of *Ca.* M. alvus (Mx05 and Mx1201), *Ca.* Methanomassiliicoccus intestinalis and *M.*

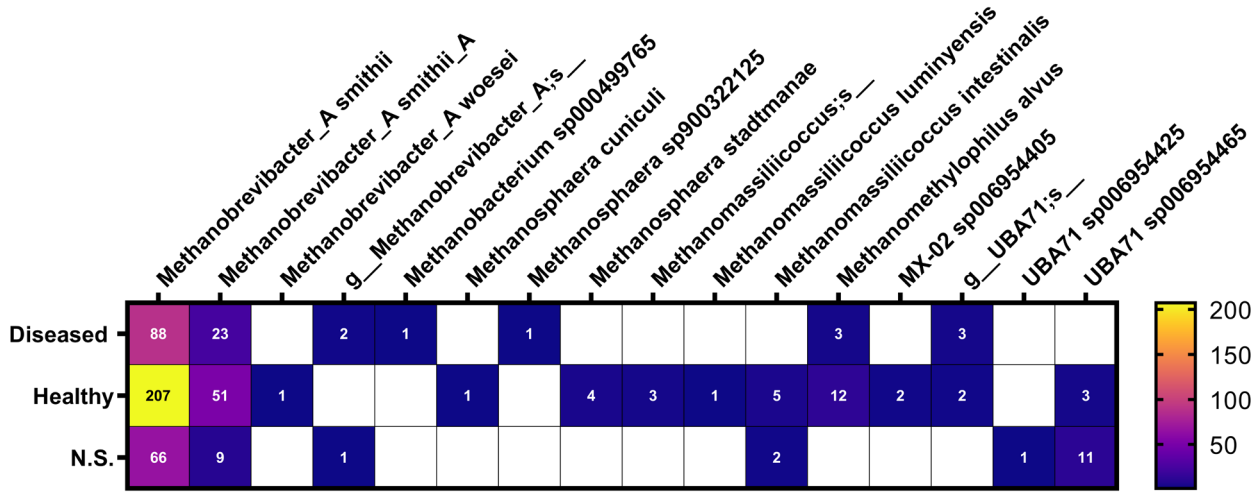


Figure 3.4. Heatmap displaying the health status distribution of human-associated methanogen MAGs and isolate genomes. Taxonomic classification of methanogen MAGs and isolate genomes was conducted using GTDB-tk (v1.3.0) and visualised using GraphPad Prism 9. Health status of the MGS used to recovery MAGs and isolate genomes was assigned according to available metadata, as per the legend. N.S., Not specified.

luminyensis were only recovered from French individuals. Comparatively, only two cultured isolates were available for *M. stadtmanae*; *M. stadtmanae* DSMZ3091 recovered from an individual from the USA and *M. stadtmanae* PA5 recently recovered from a healthy Australian faecal sample (Hoedt et al., 2018).

3.3.3 Health status distribution of human-associated methanogen MAGs and isolate genomes

Of the recovered MAGs and isolate genomes, 121 (24%) originated from samples with a recorded health disorder or disease, and 292 (58%) were identified as healthy subjects (Figure 3.4). As shown in Figure 3.5, the MAGs recovered from diseased samples were spread across most of the identified methanogen lineages. All recovered representatives of *Methanobacterium*, *M. cuniculi*, *M. stadtmanae,* uncharacterised *Methanomassiliicoccus*, *M. luminyensis*, and Methanogenic archaea Mx-02 were recovered from healthy individuals. Regarding the human *Methanobrevibacter,* MAGs recovered from healthy and diseased individuals did not contain an increase of *M. smithii* or *M. smithii*_A, with the latter comprising ~30% of each group (Figure 3.4). However, it is worth noting *M. smithii* contained a larger number of genomes without a specified health status due to limited available metadata. The representative of *Methanosphaera* sp. RUG761 (GTDB-tk classification sp900322125) was recovered from a French adult diagnosed with colorectal cancer. Interestingly, genomes classified to the *Methanomethylophilaceae* genus UBA71 contained a higher percentage of genomes derived from diseased samples (57%), although only seven genomes were assigned to this lineage.

Based on the wide distribution of methanogen lineages recovered from diseased individuals, I decided to look at the potential enrichment of *M. smithii* phylogroups recovered from UC and CD samples, given associations observed between methanogens and IBD (Blais Lecours et al., 2014; Scanlan et al., 2008). Additionally, research conducted by our collaborators suggested patients with type 2 diabetes (T2D) show altered breath methane production (data not shown) and thus MAGs recovered from patients with T2D were also included in the analyses. Interestingly, all *Methanobrevibacter* recovered from CD samples were classified as *M. smithii*_A (n=3) compared to MAGs recovered from patients with UC which were assigned to both *M. smithii* and *M. smithii*_A (n=22)(Figure 3.5). Comparatively, T2D MAGs were specifically enriched for *M. smithii* (n=9)(Figure 3.5). Although the number of genomes for each disease was limited, specifically in the case of CD, the enrichment of specific lineages was shown for these disease states.

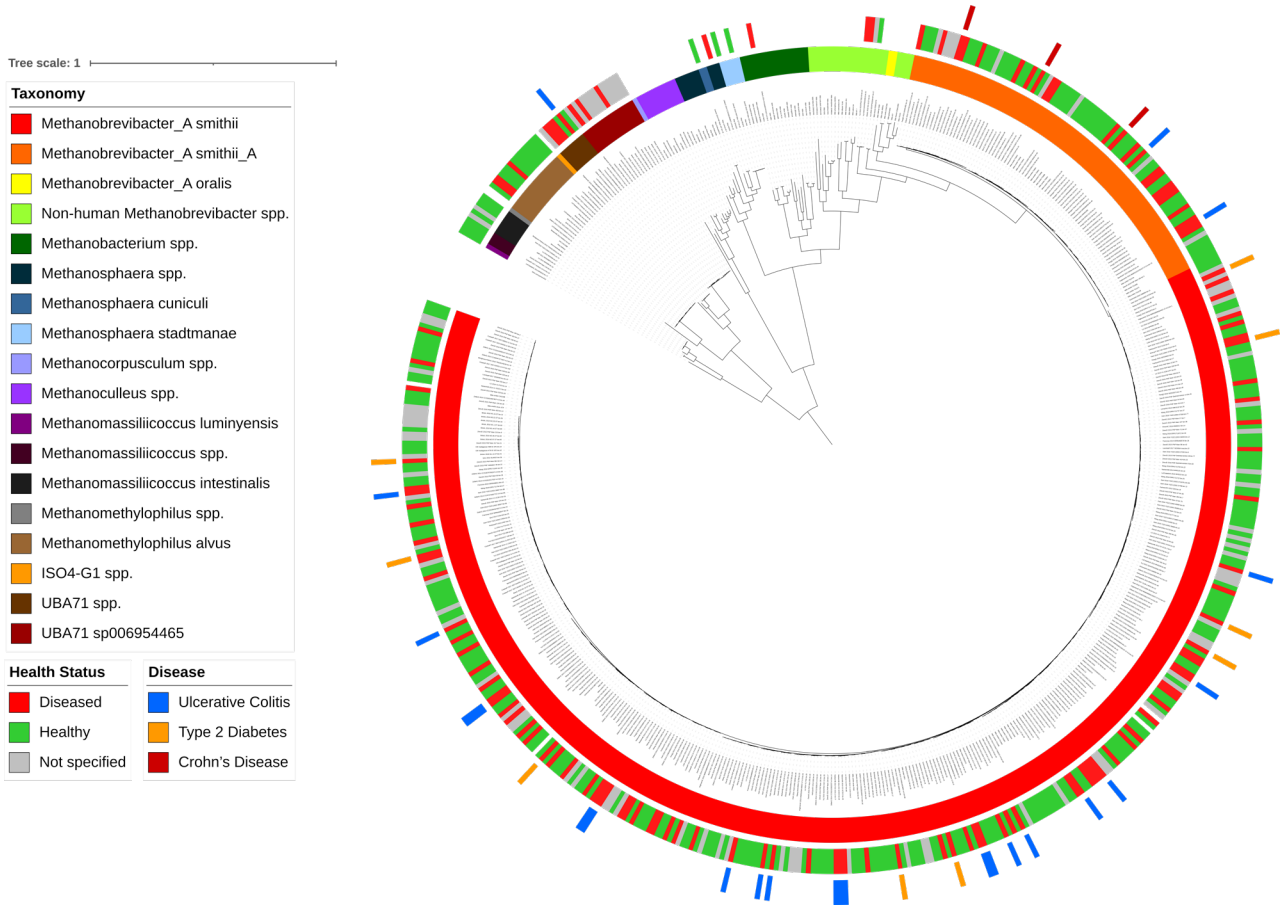


Figure 3.5. Health status distribution of human-associated methanogen MAGs and isolate genomes. The phylogeny of human-derived methanogen MAGs and reference genomes was analysed using GTDB-tk (v1.3.0), FastTree (v2.1.10), and visualised using iToL (https://itol.embl.de/). The *Methanomassiliicoccales*-associated clade was chosen as the outgroup. Health status of the MGS used to recovery MAGs and isolate genomes was assigned according to available metadata, as per the legend. The disease statuses include: Obesity, Type 2 Diabetes (T2D), Colorectal Cancer, Normal Glucose Tolerance, Impaired Glucose Tolerance, Small Adenoma, Large Adenoma, Carcinoma, Shiga-toxigenic Escherichia coli, non-alcoholic fatty liver disease, *Clostridioides difficile* infection, Advanced Adenoma, Ulcerative Colitis (UC), Crohn's Disease (CD), Ankylosing Spondylitis. Genomes recovered from UC, CD and T2D are highlighted separately, as per the legend.

Comparative analysis of the *M. smithii* disease groups with EnrichM showed a significant differential enrichment in KOs between CD and T2D genomes and CD and UC genomes (Figure 3.6-3.7). Statistical comparisons between T2D and UC genomes showed no significant enrichment of KOs. However, 12 KOs were differentially enriched between T2D and CD, with nine enriched in T2D genomes and three in CD genomes. The T2D genomes were significantly enriched for membrane transport genes, specifically components of molybdate transport system proteins (*modABC*) involved

in the uptake of molybdate (Figure 3.6)(Self et al., 2001). The T2D genomes also contained a gene annotated as 3',5'-cyclic-AMP phosphodiesterase (*cpdA*) involved in purine metabolism and biofilm formation. Comparatively, the CD genomes were differentially enriched for UDP-glucose/galactose:(glucosyl)LPS alpha-1,2-glucosyl/galactosyltransferase (*waaR, waaT, rfaJ*)

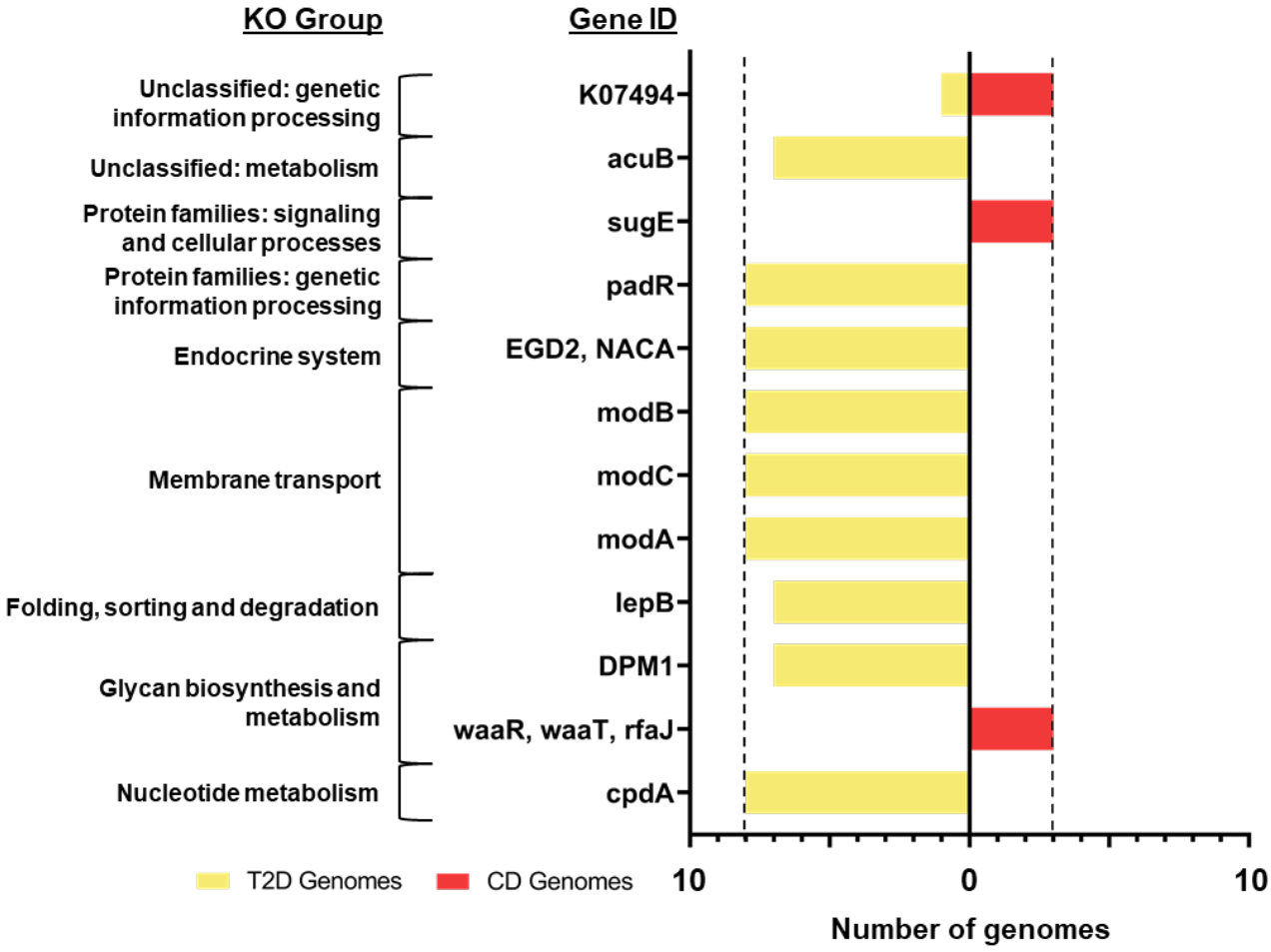


Figure 3.6. KEGG Orthologs differentially enriched in *M. smithii* MAGs recovered from T2D and CD MGS. KO annotation and statistical analysis was performed using the 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). KOs with p values of <0.05 were retained, as determined by Fisher's Exact Test. Three CD and eight T2D *M. smithii* MAGs were included in the analysis. The full list of KOs is shown in Table 6.6.

involved in membrane glycosylation. Additionally, the CD genomes were also enriched for quaternary ammonium compound-resistance protein SugE which provides resistance to quaternary ammonium compounds. Acetoin utilisation protein AcuB may also allow for the differential use of acetoin as a carbon source in the T2D genomes.

Compared with the UC genomes, SugE and K03276 were again enriched in the CD genomes, along with a putative transposase (K07494)(Figure 3.7). Interestingly, the UC genomes contained all

enriched genes shown in the comparison between the T2D and CD genomes (Figure 3.6-3.7). However, the UC genomes were also enriched for signal peptidase I (*lepB*) and an HSP20 family protein involved in folding, sorting and degradation (Figure 3.7). Additionally, the UC genomes were also enriched for nucleolar GTP-binding protein *NOG1* involved in translation. Interestingly, the UC and CD genomes were differentially enriched for glycan biosynthesis and metabolism, with UC genomes enriched for dolichol-phosphate mannosyltransferase (*DPM1*) and CD genomes enriched for UDP-glucose/galactose:(glucosyl) LPS alpha-1,2-glucosyl/galactosyltransferase (*waaR, waaT, rfaJ*). Genes annotated as cazymes showed no significant enrichment between T2D and UC or T2D and CD genomes. In comparison to UC, the CD genomes showed significant enrichment of genes annotated as GT2 and GT8 (P=0.0119, P=0.0198), though the corrected p-values did not maintain significance due to the small number of genomes in the CD group (P= 0.1087, P= 0.1087).

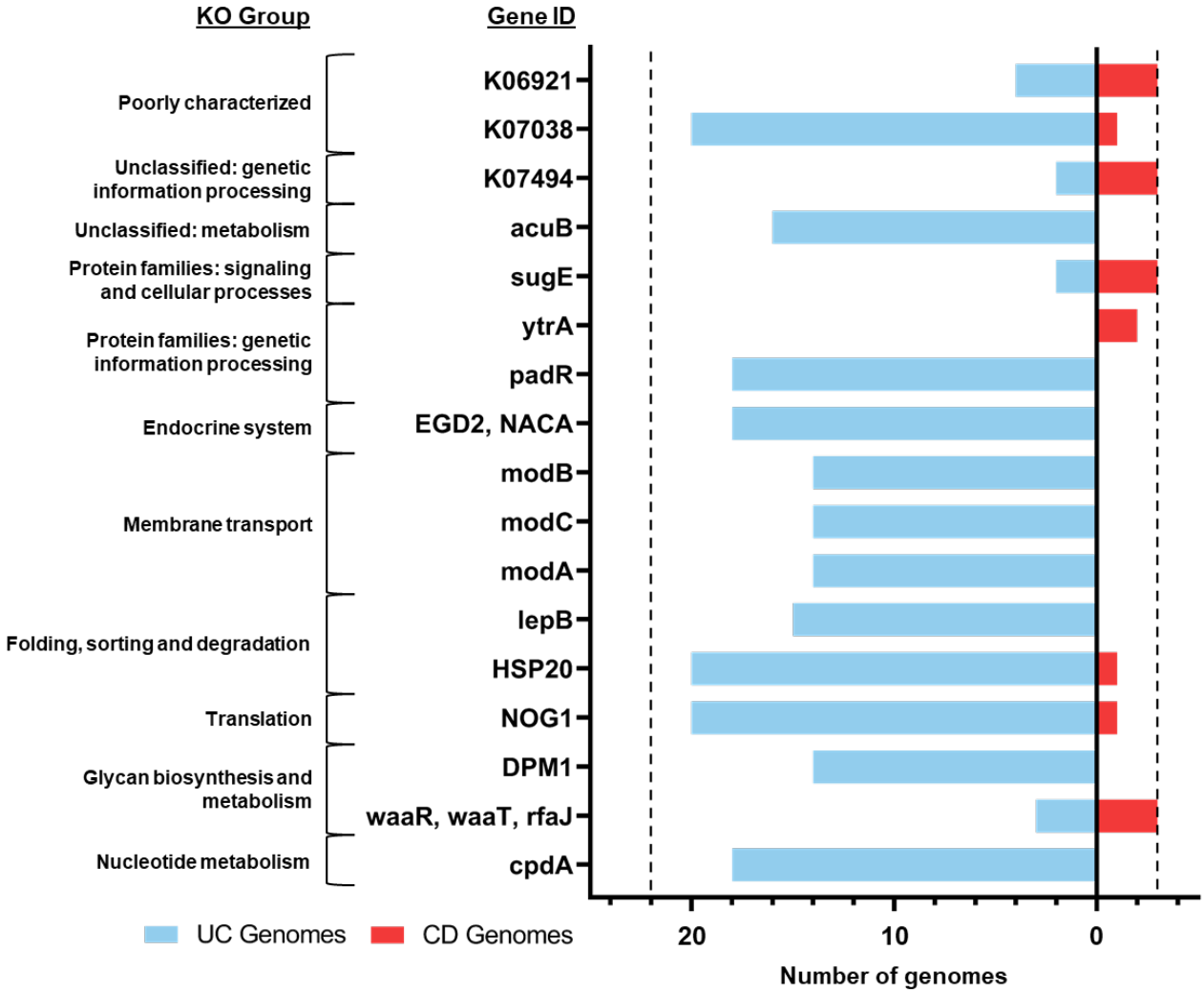


Figure 3.7. KEGG Orthologs differentially enriched in *M. smithii* MAGs recovered from UC and CD MGS. KO annotation and statistical analysis was performed using the 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). KOs with p values of <0.05 were retained, as determined by Fisher's Exact Test. Three CD and 22 UC *M. smithii* MAGs were included in the analysis. The full list of KOs is shown in Table 6.7.

3.3.4 The *M. smithii* and *M. smithii*_A phylogroups show distinct genetic differences

Historically, *M. smithii* was considered a single phylogroup with relatively consistent genetic potential. However, as shown in Section 3.3.1, *M. smithii*_A appears to be a phylogenetically distinct subgroup of *M. smithii* that is yet to be well defined. Here I investigate the two groups to determine genetic differences between these newly defined *M. smithii* phylogroups.

As a part of the initial analyses, basic genomic statistics were used to compare between the *M. smithii* and *M. smithii*_A phylogroups (Table 3.3). Comparisons of genome quality showed no significant difference in the completeness or contamination scores between the two groups. Additionally, no difference was observed between the number of contigs, however, *M. smithii*_A showed a significantly smaller (N50) contig length (p=0.0307). *M. smithii*_A also contained a significantly larger genome size and a greater number of predicted coding sequences compared to *M. smithii* (p≤0.0001; p≤0.0001), suggesting *M. smithii*_A contains additional genetic elements and coding potential (Table 3.3). This statistical difference was also maintained when accounting for average genome completeness.

Table 3.3. Average genome statistics of *M. smithii* and *M. smithii*_A. Average genomes values were calculated using CheckM (v1.0.7). Statistical comparisons were conducted using Student's T Test in GraphPad Prism 9. The predicted genome size and predicted CDS was included to account for genome completeness. A p-value ≤0.05 was considered significant. CDS, coding sequences.

	M. smithii	***M. smithii*_A**	**P value**
No. genomes	369	84	-
No. cultured isolates	13	14	-
Genome size	1726729 (±114728)	1842210 (±110698)	<0.0001
Predicted genome size	1752851 (±111370)	1865864 (±103482)	<0.0001
CDS	1724 (±139)	1818 (±115)	<0.0001
Predicted CDS	1751 (±149)	1843 (±122)	<0.0001
Contigs	82 (±98)	93 (±92)	0.3483
N50 (contigs)	117642 (±186067)	72843 (±58143)	0.0307
Completeness	98.51 (±2.43)	98.72 (±1.96)	0.4603
Contamination	0.31 (±0.77)	0.46 (±1.00)	0.1296

Using EnrichM, *M. smithii* and *M. smithii*_A showed separation based on gene orthologs, as shown in Figure 3.8, though it is worth noting the low variance for PC1 and PC2. Separation was also shown for the *M. woesei* and *M. woesei*-like groups of *Methanobrevibacter* genomes. Interestingly, MAGs recovered from non-western populations showed the greatest variance, as shown by the labelled MAGs recovered from native American and Chinese cohorts (Figure 3.8).

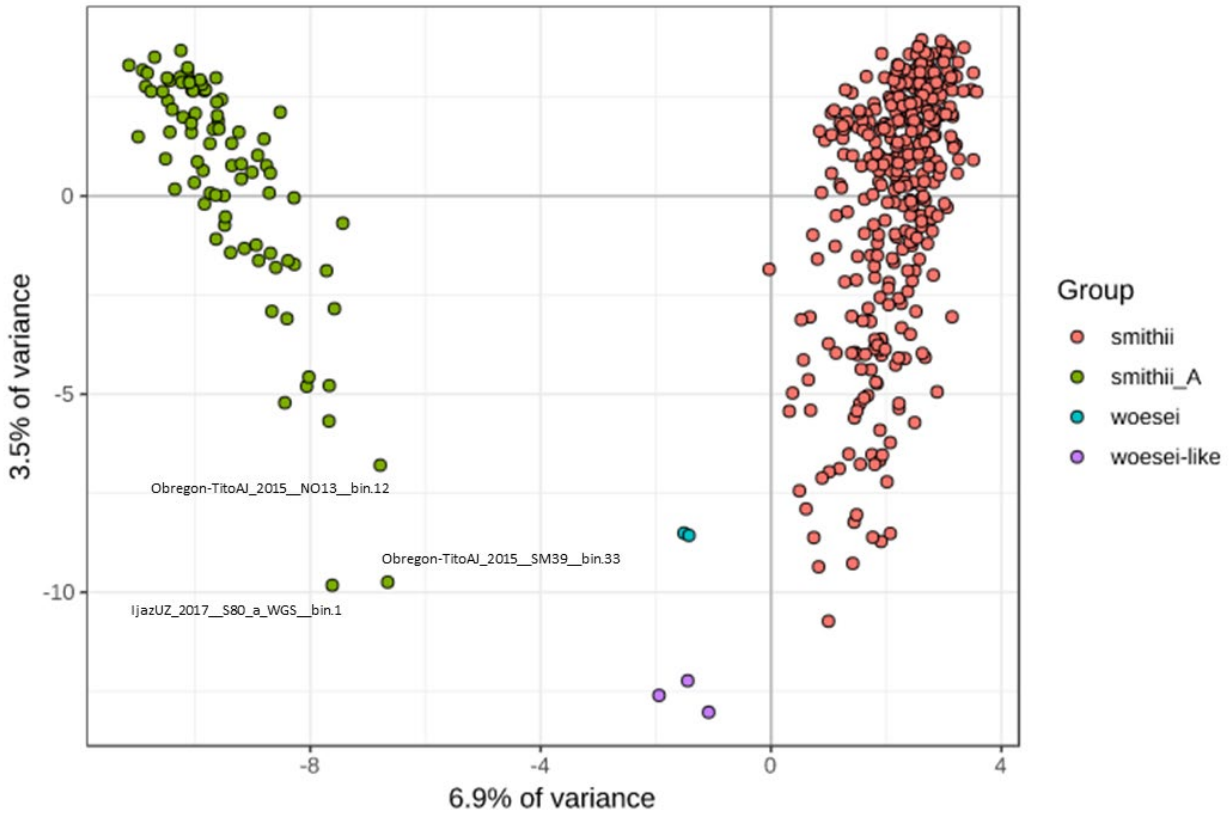


Figure 3.8. Principle component analysis (PCA) showing the variance between *Methanobrevibacter* genomes according to gene orthologs. Ortholog annotation, statistical analysis and PCA plot generation was performed using EnrichM (v0.4.9). Phylogenetic groups are shown with different coloured circles, according to the legend. A clear separation is shown between the *M. smithii* and *M. smithii*_A subgroups according to gene orthologs.

*M. smithii*_A was significantly enriched for GT8 with Fisher's Exact test and Mann-Whitney U tests (corrected p=8.93E-38; corrected p=4.03E-38), with 92.8% of *M. smithii*_A genomes containing a GT8 gene annotation compared to 15.8% of *M. smithii*. GT8 is likely associated with lipopolysaccharide glycosylation. GH99 also seemed to be enriched in *M. smithii*_A but did not reach significance with p-value correction. Together these annotations suggest *M. smithii* and *M. smithii*_A are differentially enriched for genes associated with glycosylation, which could result in differential expression of outer membrane structures and surface structure glycosylation.

Annotation of genes with KOs showed 34 differentially enriched genes between *M. smithii* or *M. smithii*_A (Figure 3.9). The largest category of differentially enriched genes was signalling and cellular processes, as well as membrane transport. *M. smithii*_A was enriched for iron complex transport system permease protein ABC.FEV.P (K02015), iron complex transport system substrate-binding protein ABC.FEV.S (K02016), and iron complex transport system ATP-binding protein ABC.FEV.A (K02013)(figure 3.10B). Although, 20% of *M. smithii* genomes also contained the

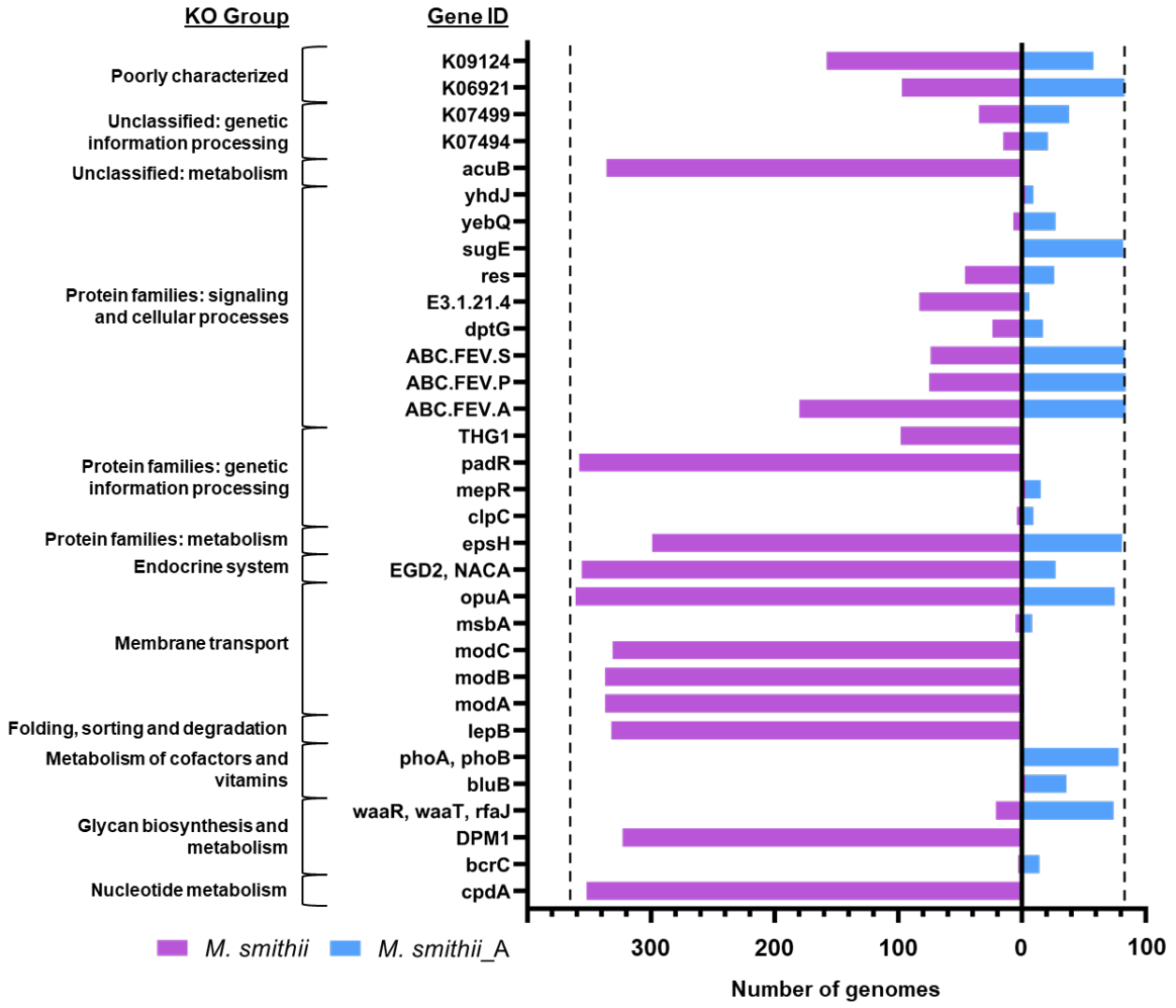


Figure 3.9. KEGG Orthologs differentially enriched in *M. smithii* and *M. smithii*_A. KO annotation and statistical analysis was performed using the 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). KOs with a corrected p value of <0.05 were retained, as determined by Fisher's Exact Test. 84 *M. smithii*_A and 369 *M. smithii* were included in the analysis. The full list of KOs is available in Table 6.8.

ABC.FEV.P and *ABC.FEV.S,* and 49% contained the *ABC.FEV.A* gene. Only two genomes between *M. smithii* and *M. smithii*_A contained annotated iron complex outer membrane receptor protein TC.FEV.OM (K02014). Comparatively, *M. smithii* showed enrichment of molybdate transport system substrate-binding protein *modA* (K02020), molybdate transport system permease protein *modB* (K02018), molybdate transport system ATP-binding protein *modC* (K02017), as displayed in Figure 3.10A. Together, this suggests the phylogroups of *M. smithii* possess a differential requirement of trace metals, from which *M. smithii* had adapted for increased uptake of molybdate and *M. smithii*_A for the uptake of iron. In addition to metal uptake, *M. smithii*_A was enriched for MFS multidrug resistance protein *yebQ* and quaternary ammonium compound-resistance protein *SugE*

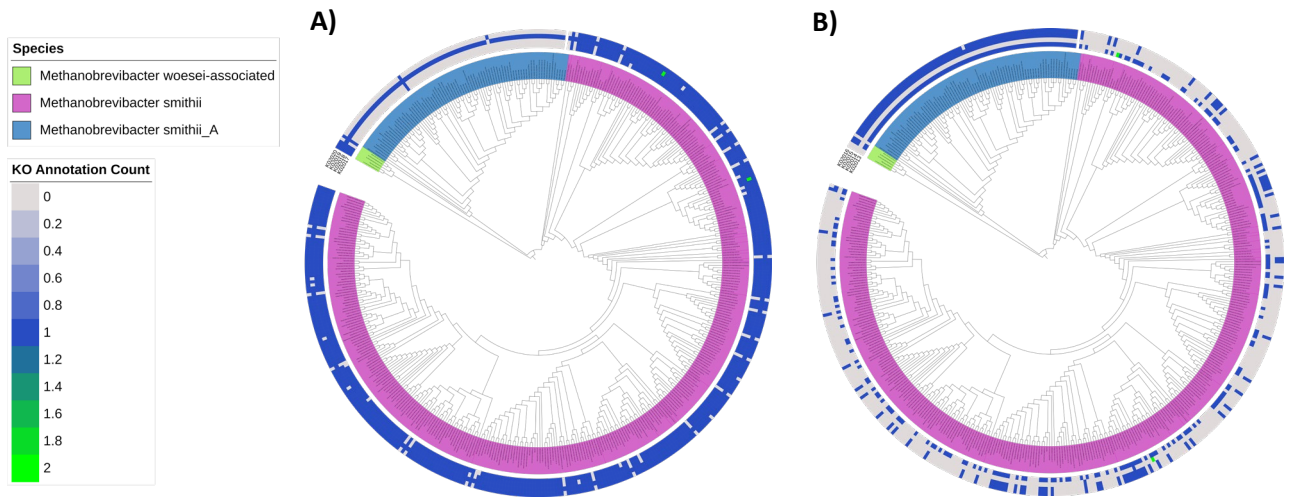


Figure 3.10. Phylogenetic analysis of human *Methanobrevibacter* with annotated metal uptake genes. Phylogeny of *Methanobrevibacter* genomes were analysed using GTDB-tk (v1.3.0), FastTree (v2.1.10), and visualised using iToL (https://itol.embl.de/). Different *M. smithii* phylogroups are shown in purple and blue, and *M. woesei*-like are shown in green. The *M. woesei*-like clade was chosen as the outgroup. A) show molybdate transport system proteins (K02017, K02018, K02019, K02020) enriched in *M. smithii*. B) shows iron complex transport proteins (K02013, K02014, K02015, K02016) enriched in the currently unrecognised *M. smithii*_A.

(Figure 3.9). In terms of prokaryotic defence, *M. smithii*_A was also enriched for type II restriction enzyme (E3.1.21.4), type III restriction enzyme (*res*), adenine-specific DNA-methyltransferase (*yhdJ*), and DNA phosphorothioation-dependent restriction protein DptG (Figure 3.9).

M. smithii showed significant enrichment genetic information and processing with PadR family transcriptional regulator gene and tRNA(His) guanylyltransferase (*THG1*). Additionally, acetoin utilisation protein AcuB was also enriched and may allow for the *M. smithii* to use a wider range of carbon sources than *M. smithii*_A. Interestingly, *M. smithii*_A was enriched for a gene involved in the metabolism of cofactors and vitamins, specially 5,6-dimethylbenzimidazole synthase (*bluB*) involved in riboflavin metabolism and alkaline phosphatase (*phoA/phoB*) involved in thiamine metabolism.

Another key category of differentially enriched genes was those involved in glycan biosynthesis and metabolism. *M. smithii*_A was specifically enriched for UDP-glucose/galactose:(glucosyl)LPS alpha-1,2-glucosyl/galactosyltransferase (*waaR, waaT, rfaJ*) involved in the glycosylation of outer membrane lipopolysaccharide, as well as undecaprenyl-diphosphatase (*bcrC*), which contributes to peptidoglycan biosynthesis. Glycosyltransferase *epsH* is also enriched in *M. smithii*_A, though the

majority of *M. smithii* genomes also encode this gene. Comparatively, dolichol-phosphate mannosyltransferase (*DPM1*) is significantly enriched in *M. smithii* and not found in *M. smithii*_A.

Collectively, these analyses show that *M. smithii* and *M. smithii*_A encode for a subset of differential genes unique to the respective phylogroups, and involved in membrane transport, glycan biosynthesis and carbohydrate metabolsm. These findings provide a genetic basis that further validates the phylogenetic separation of *M. smithii* and *M. smithii*_A as unique phylogroups.

3.3.5 Differential enrichment of genes within subgroups of *M. smithii* and *M. smithii*_A

M. smithii was separated into 11 distinct subgroups containing at least 10 genomes, accounting for more than 90% of *M. smithii* genomes (Figure 3.11). Analysis of cazymes showed a slight variation within each subgroup, though no significant enrichment was observed in any *M. smithii* subgroup. However, a comparison of KO genes enriched between subgroups showed a clear separation in the *M. smithii* subgroups (see Table 6.9 for the complete list of enriched KOs). Subgroup 5 showed the greatest average number of differentially enriched KOs at 13.9, suggesting this subgroup was the most dissimilar on average to all other subgroups, although only three genes were enriched between subgroups 5 and 3. Interestingly, this subgroup contained the cultured representative of *M. smithii* PS, the type strain of *M. smithii* (Figure 3.11). Similarly, subgroup 11 also showed many differentially enriched genes compared to other subgroups but did not contain a cultured representative. Subgroups 4 and 9 showed the fewest number of average enriched KOs at 4.2 and 5, respectively. As such, *M. smithii* DSMZ3275 may be a more appropriate cultured representative for the 'average' *M. smithii* compared to *M. smithii* PS.

In terms of specific KOs, CRISPR-associated genes, restriction enzymes, and transposases account for 20.5% of KOs enriched between different subgroups, such as subgroup 7, which was enriched for *cas1*, *csh1*, *csh2*, *csm1*, *csm4*, *csm5*, and *cas5h* compared to subgroup 1 (Table 6.10). As discussed in Section 3.3.4, *M. smithii*_A was enriched for iron complex transport systems but specific subgroups of *M. smithii* were also enriched for these genes. Specifically, subgroups 10, 5, 6, and 9 were enriched for *ABC.FEV.S*, *ABC.FEV.P* and *ABC.FEV.A*, compared to the other subgroups. Two annotations of alcohol dehydrogenase genes are also differentially enriched: *AKR1A1* alcohol dehydrogenase (NADP+)(K00002) and *adh1* alcohol dehydrogenase (K19954). *AKR1A1* was encoded in all subgroups, except for subgroup 5, in which only 67% of the genomes contained the gene. Interestingly, *adh1* was only found in 16 genomes within subgroup 11 (35.5%). Subgroup 11 was also enriched for phosphonopyruvate decarboxylase and phosphoenolpyruvate phosphomutase, both involved in phosphonic acid metabolism (Yu et al., 2013). Glycosyltransferases were also

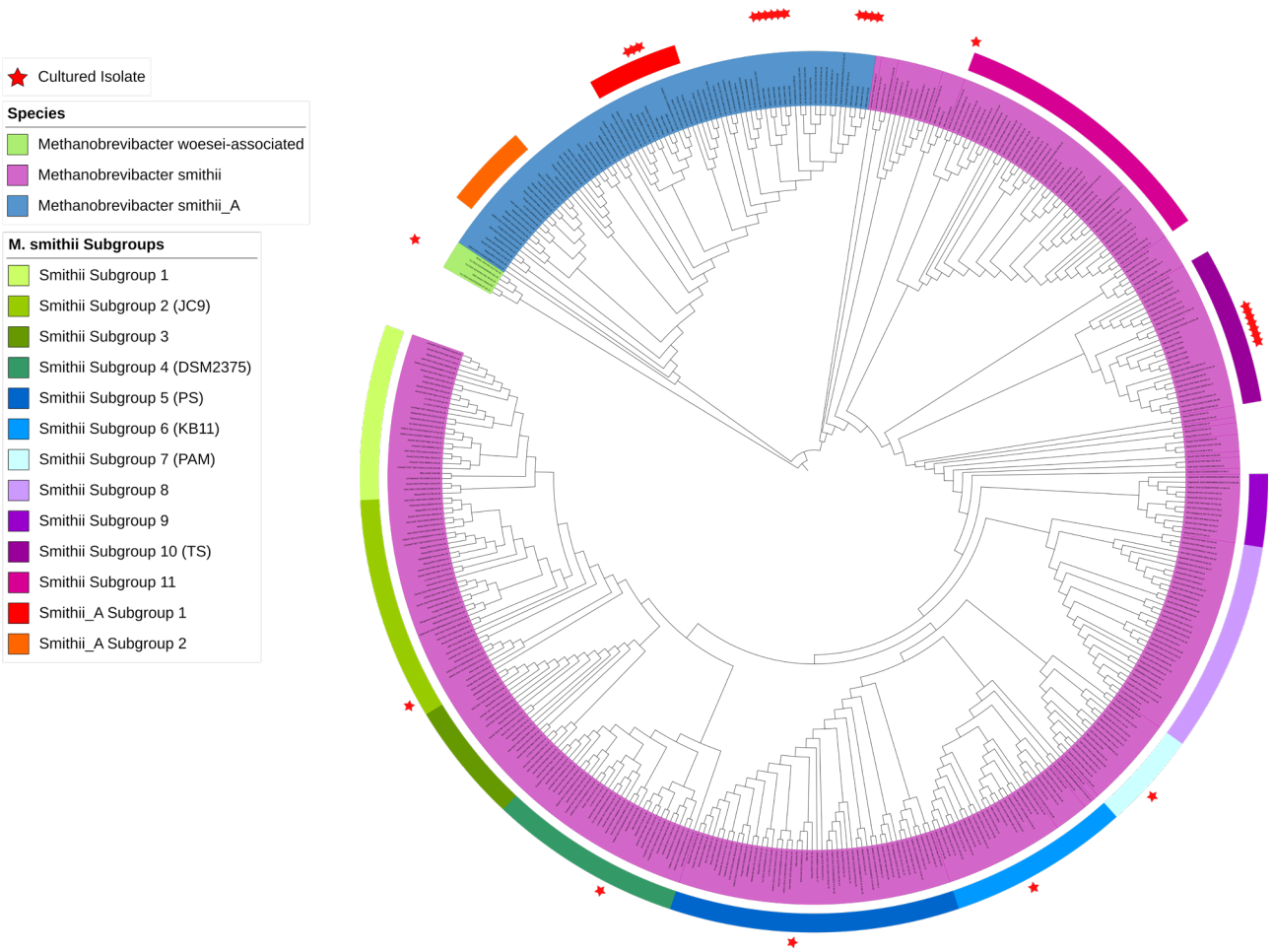


Figure 3.11. Identification of major *M. smithii* and *M. smithii*_A subgroups. Phylogeny of *Methanobrevibacter* genomes were analysed using GTDB-tk (v1.3.0), FastTree (v2.1.10) and visualised using iToL (https://itol.embl.de/). The *Methanobrevibacter woesei*-associated clade was chosen as the outgroup. Major subgroups of *M. smithii* and *M. smithii*_A containing at least 10 genomes were identified according to the legend. Cultured isolates are identified by the red stars.

differentially enriched, specifically glycosyltransferase *epsJ*, glycosyltransferase *epsH*, and 1,2-diacylglycerol 3-beta-glucosyltransferase involved in glycerolipid metabolism, as well as dolichol-phosphate mannosyltransferase. It is worth noting that only phosphoenolpyruvate phosphomutase, phosphonopyruvate decarboxylase, and *adh1* of subgroup 11 maintained significance with p-value correction compared with subgroups 1, 2, 4, 5, and 8.

Compared to *M. smithii*, *M. smithii*_A appears to consist of several smaller clusters of genomes, with only two relatively large subgroups (Figure 3.11). Again, like the *M. smithii* subgroups, *M. smithii*_A genomes showed a small number of cazymes variations between the subgroups, although these differentiations showed no statistical significance. However, several KO annotations did appear to show enrichment in the subgroups of *M. smithii*_A (Table 3.4). CRISPR-associated proteins *csm1-5* were enriched in *smithii*_A subgroup 2, which remained consistent with analyses by Fisher's Exact

and Mann-Whitney U Tests. Additionally, analyses by Mann-Whitney U test also showed an enrichment of *pflACE*; pyruvate formate lyase-activating enzyme (EC:1.97.1.4), DNA mismatch endonuclease and patch repair protein *vsr*. Comparatively, *smithii*_A subgroup 1 was enriched for a putative transposase (K07494). The Mann-Whitney U test also showed an enrichment of the F420-non-reducing hydrogenase iron-sulfur subunit (*mvhD, vhuD, vhcD*), the formate dehydrogenase

Table 3.4. KEGG Orthology gene annotations enriched in *M. smithii*_A subgroups. EnrichM (v0.4.9) was used to determined KOs enriched between *M. smithii*_A subgroup 1 (n=15) and *M. smithii*_A subgroup 2 (n=14). KO annotations with p-values ≤0.05 were included. Due to the small number of genomes in each group, no KO maintained significance with p-value correction. Fisher's Exact test (top) and Mann-Whitney U test (bottom) were used for statistical comparison of genomes containing KOs and mean gene counts, respectively.

Fisher's Exact test				
KO	Subgroup 2	Subgroup 1	P value	Gene Description
K07494	2	11	0.0025	K07494; putative transposase
K04096	6	0	0.0063	smf; DNA processing protein
K19138	10	3	0.0092	csm2; CRISPR-associated protein Csm2
K09002	10	3	0.0092	csm3; CRISPR-associated protein Csm3
K19140	10	3	0.0092	csm5; CRISPR-associated protein Csm5
K19139	9	3	0.0253	csm4; CRISPR-associated protein Csm4
K07016	9	3	0.0253	csm1, cas10; CRISPR-associated protein Csm1
Mann-Whitney U test				
KO	Subgroup 2	Subgroup 1	P value	Gene Description
K07494	0.1429	0.7333	0.0009	K07494; putative transposase
K14127	1.2143	1.7333	0.0033	mvhD, vhuD, vhcD; F420-non-reducing hydrogenase iron-sulfur subunit
K19138	0.7143	0.2000	0.0034	csm2; CRISPR-associated protein Csm2
K09002	0.7143	0.2000	0.0034	csm3; CRISPR-associated protein Csm3
K19140	0.7143	0.2000	0.0034	csm5; CRISPR-associated protein Csm5
K00123	1.2143	1.6667	0.0087	fdoG, fdhF, fdwA; formate dehydrogenase major subunit [EC:1.17.1.9]
K00125	1.2143	1.6667	0.0087	fdhB; formate dehydrogenase (coenzyme F420) beta subunit
K19139	0.6429	0.2000	0.0093	csm4; CRISPR-associated protein Csm4
K07016	0.6429	0.2000	0.0093	csm1, cas10; CRISPR-associated protein Csm1
K09124	0.5000	0.8667	0.0193	K09124; uncharacterized protein
K04069	2.5000	2.1333	0.0225	pflA, pflC, pflE; pyruvate formate lyase activating enzyme [EC:1.97.1.4]
K00012	0.6429	0.9333	0.0311	UGDH, ugd; UDPglucose 6-dehydrogenase [EC:1.1.1.22]
K09131	0.6429	0.9333	0.0311	K09131; uncharacterized protein
K07458	0.4286	0.1333	0.0429	vsr; DNA mismatch endonuclease, patch repair protein [EC:3.1.-.-]
K03320	0.8571	1.0667	0.0452	amt, AMT, MEP; ammonium transporter, Amt family
K04751	0.8571	1.0667	0.0452	glnB; nitrogen regulatory protein P-II 1
K03606	1.0714	0.8667	0.0493	wcaJ; putative colanic acid biosysnthesis UDP-glucose lipid carrier transferase
K06147	2.0714	1.8667	0.0493	ABCB-BAC; ATP-binding cassette, subfamily B, bacterial

major subunit (*fdoG, fdhF, fdwA*), and formate dehydrogenase (coenzyme F420) beta subunit (*fdhB*). These are all potentially major genes involved in methanogenesis and energy production, specifically the utilisation of formate by formate dehydrogenases (Wood et al., 2003). Similarly, ammonium transporter family protein AMT and nitrogen regulatory protein *glnB* were also enriched in subgroup 2. Given the small number of genomes in each subgroup (*Smithii*_A subgroup 1 n=15, *Smithii*_A subgroup 2 n=14), these KOs represent potentially differentially enriched genes between the different subgroups. A larger number of genomes for each subgroup is required for statistical confirmation of the differentially enriched genes with p-value correction.

3.3.6 Recovery of novel lineages of human-associated *Methanobrevibacter*, *Methanosphaera*, and *Methanobacterium*

As a part of the recovery of archaeal MAGs from human metagenomes by Pasolli et al. (2019), several MAGs were recovered that represented novel lineages of human-associated methanogens (Table 3.5). One *Methanobrevibacter* MAG closely aligned with *Methanobrevibacter woesei*, originally isolated from goose faeces (Miller et al., 1986). Three additional species of *Methanobrevibacter* clustered separately from all other genomes. Blastn analysis of the 16S rRNA gene of YuJ_2015_SZAXPI017581-93_bin.23 showed the highest sequence similarity (99.64%) to Uncultured archaeon clone 2E5 (HQ678043.1) recovered from a low-temperature anaerobic bioreactor inoculated with pig manure. The sequence also shared 97.24% sequenced similarity to *M. smithii* PS, suggesting this MAG is a closely related but distinct clade of *Methanobrevibacter* that may also be found in pigs. Interestingly, these novel lineages of *Methanobrevibacter* were only recovered from two studies, with three of the genomes being recovered from a single study by Yu et al. (2017) on the microbiome of CRC patients. The genomes were recovered from three Chinese

Table 3.5. Genomic features of human-derived HQ MAGs associated with novel methanogen lineages. MAGs were recovered by Pasolli et al. (2019) and taxonomically classified using GTDB-tk (v1.3.0). Quality assessment and basic genomic details were determined using CheckM (v1.0.7).

Genome	Classification	Completeness	Contamination	No. tRNAs (/20)	Genome size	Contigs	Predicted genes
YuJ_2015_SZAXPI015264-87_bin.27	Methanobrevibacter_A woesei	100	0	19	1610769	14	1622
BritoIL_2016_W2.43.ST_bin.33	Methanobrevibacter_A;s_	99.2	0	18	1691878	132	1764
YuJ_2015_SZAXPI017581-93_bin.23	Methanobrevibacter_A;s_	98.74	0	11	1519580	241	1543
YuJ_2015_SZAXPI003409-8_bin.32	Methanobrevibacter_A;s_	97.6	0.8	18	1601068	76	1674
ZellerG_2014_CCIS33816588ST-4-0_bin.32	Methanosphaera sp900322125	97.6	0	18	1759917	90	1657
ZellerG_2014_CCIS46047672ST-4-0_bin.2	Methanobacterium sp000499765	98.13	0	19	1905101	37	1916
ZeeviD_2015_PNP_Main_437_bin.13	Methanosphaera cuniculi	95.33	0.8	19	1661817	245	1512

individuals, one healthy and two with CRC. The last genome was recovered from a study on bacterial transmission of a Fijian population (Brito et al., 2019).

In addition to the *Methanobrevibacter*, novel lineages of *Methanosphaera* were also recovered. One HQ MAG was classified as a representative of *M. cuniculi*, a species of *Methanosphaera* originally isolated from a rabbit (Biavati et al., 1988). In fact, the McrA protein sequence of ZeeviD_2015_PNP_Main_437_bin.13 showed 100% amino acid sequence identity to *M. cuniculi.* This MAG was recovered from a study on nutritional intervention to predict glycaemic responses in a population of Israeli adults (Zeevi et al., 2015). Similarly, one MAG was classified as *Methanosphaera* 'sp900322125', the GTDB-tk classification given to a cluster *Methanosphaera* MAGs containing RUG761 recovered from metagenomic samples of bovine (Stewart et al., 2019). As per my analysis in Section 3.3.1, the closest related isolate strain was *Methanosphaera* sp. WGK6,

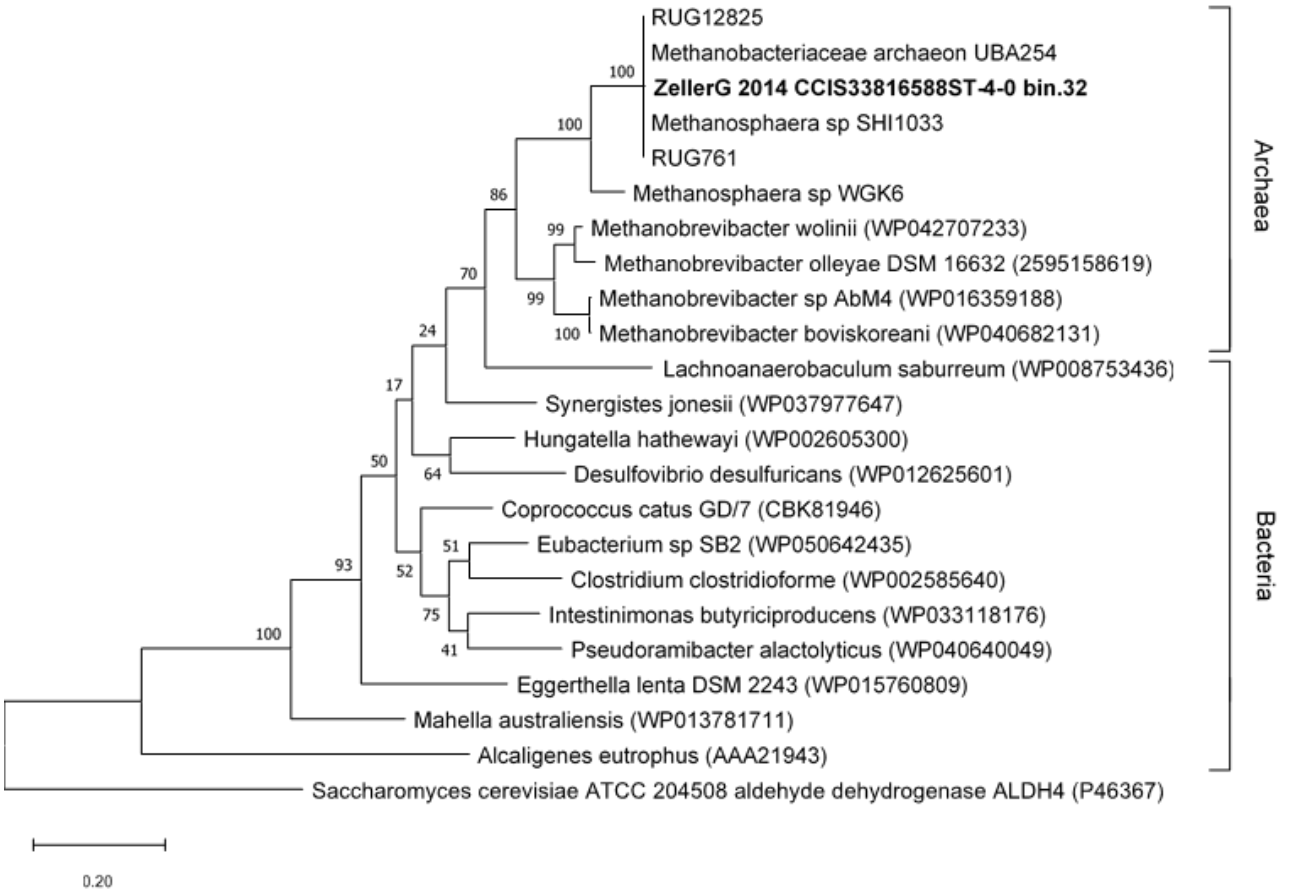


Figure 3.12. Phylogenetic analysis of predicted ethanol-utilising aldehyde dehydrogenase of *Methanosphaera*. Predicted '*Wald*' genes were recovered using Kaptive (v0.5.1) with a custom database of the '*wald*' (NL43_RS02835; WP_198923183.1) gene of *Methanosphaera* sp. WGK6. Recovered nucleotide sequences were translated to amino acid using ExPASy (https://web.expasy.org/translate/). Phylogeny was then inferred using MUSCLE in MEGA-X, and the tree constructed using Maximum-likelihood with the JJT model and 1000 bootstrap replications. *S. cerevisiae* was used to root the tree. The human derived *Methanosphaera* MAG is shown in bold.

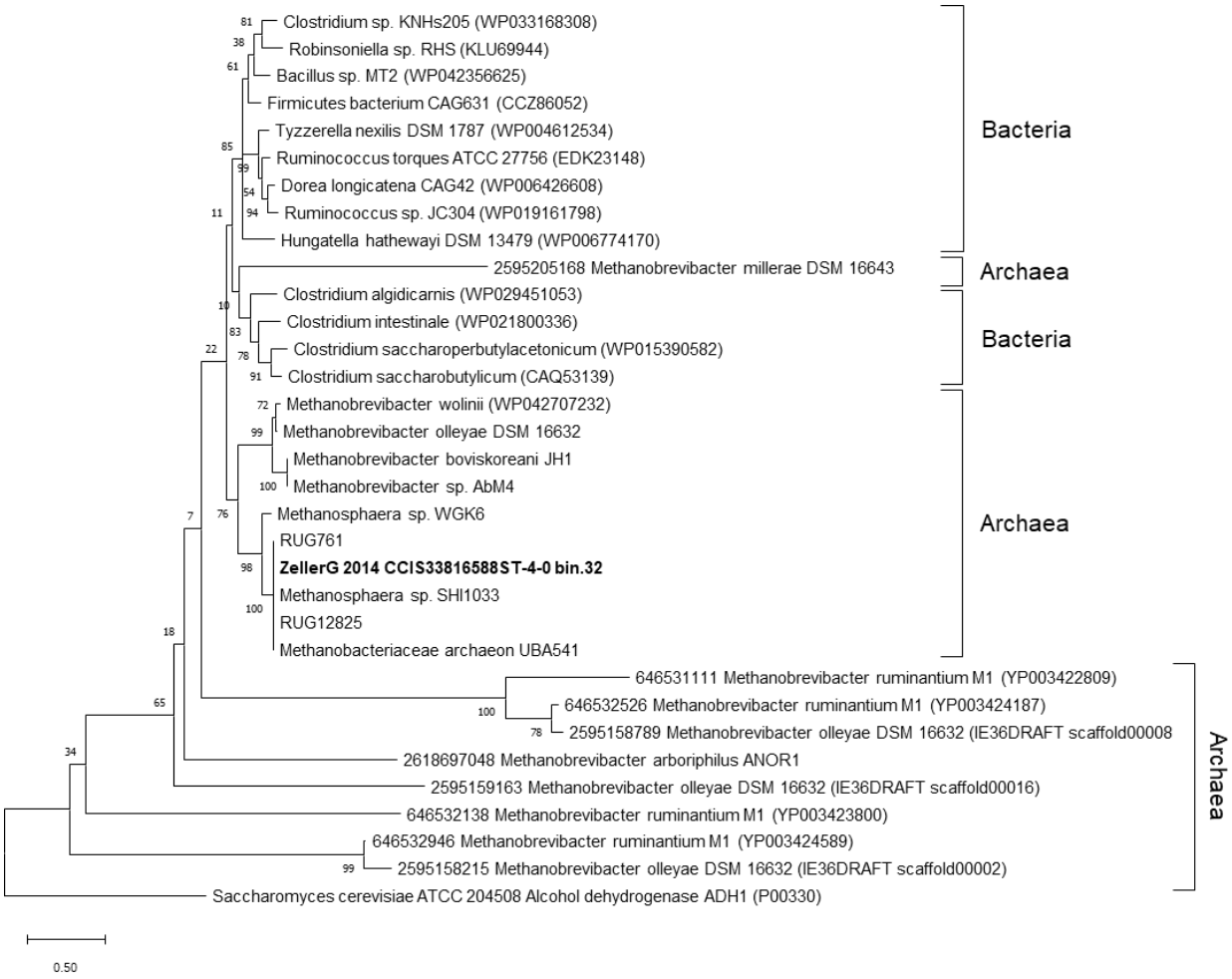


Figure 3.13. Phylogenetic analysis of predicted ethanol-utilising alcohol dehydrogenase of *Methanosphaera*. Predicted '*Walc*' genes were recovered using Kaptive (v0.5.1) with a custom database of the '*walc*' (NL43_RS02830; WP_069592539.1) gene of *Methanosphaera* sp. WGK6. Recovered nucleotide sequences were translated to amino acid using ExPASy (https://web.expasy.org/translate/). Phylogeny was then inferred using MUSCLE in MEGA-X, and the tree constructed using Maximum-likelihood with the JJT model and 1000 bootstrap replications. *S. cerevisiae* was used to root the tree. The human derived *Methanosphaera* MAG is shown in bold.

a species of *Methanosphaera* isolated from the forestomach contents of a Western grey kangaroo (Hoedt et al., 2016). This novel *Methanosphaera* MAG was recovered from a study on the early detection of CRC in patients from several countries (Zeller et al., 2014). From this study, a MAG classified as *Methanobacterium* (GTDB-tk classification sp000499765) was also recovered. Interestingly, this MAG is currently the only genome of *Methanobacterium* recovered from a human sample and closely clusters with *Methanobacterium* sp. MBI isolated from maize silage and cattle manure of a biogas plant (Maus et al., 2013). *Methanobacterium* are recognised as hydrogenotrophic methanogens, primarily utilising hydrogen and carbon dioxide for energy production (Joulian et al., 1998; Kitamura et al., 2011). However, the ZellerG_2014_CCIS46047672ST-4-0_bin.2 MAG

contains genes that may provide additional adaptability. Interestingly, two genes annotated as choline trimethylamine-lyase activating enzyme (*cutC*) are also present and catalyse the conversion of choline to trimethylamine (TMA) and acetaldehyde (Craciun & Balskus, 2012).

Both novel lineages from the Zeller et al. (2014) study and the *M. woesei*-associated MAG were recovered from individuals with CRC, suggesting potential disease-associated changes in gut ecology may facilitate conditions necessary for an increased abundance of these novel methanogen lineages. ZellerG_2014_CCIS33816588ST-4-0_bin.32 represents a lineage of *Methanosphaera* that can utilise ethanol as a substrate for methanogenesis (Hoedt et al., 2016). Using the predicted ethanol-utilising alcohol dehydrogenase '*walc*' (NL43_RS02830; WP_069592539.1) and aldehyde dehydrogenase 'wald' (NL43_RS02835; WP_198923183.1) of *Methanosphaera* sp. WGK6, ZellerG_2014_CCIS33816588ST-4-0_bin.32 was shown to contain homologous of both ethanol-utilising genes. Phylogenetic assessment of the aldehyde (Figure 3.12) and alcohol dehydrogenase (Figure 3.13) protein sequences showed both to cluster closely with the respective proteins of *Methanosphaera* sp. WGK6. However, the human *Methanosphaera* MAG clustered together with additional sequences recovered from *Methanosphaera* MAGs of sheep and bovine, suggesting this predicted utilisation of ethanol may be found in this specific group of *Methanosphaera* that colonises a wider variety of animals (Figure 3.12-3.13).

Collectively, these HQ MAGs represent the first of their respective lineages to be recovered from human samples. My analyses expands the number of methanogen lineages associated with the human microbiome and provides an insight into largely uncharacterised species and the metabolic implications they pose to the wider microbiome.

3.4 Discussion

Here I provided an in-depth analysis into human-associated methanogenic archaea derived from metagenomic datasets. A total of 55 HQ taxonomically classified archaea MAGs were successfully recovered from three publicly available human metagenomic datasets and combined with those recently produced by Pasolli et al. (2019). The human gastrointestinal microbiome is typically thought to contain three main groups of methanogens: *M. smithii*, *M. stadtmanae,* and *Methanomassiliicoccales* spp. However, among the recovered MAGs, novel lineages of *M. cuniculi*, *Methanosphaera* spp., *M. woesei, Methanobrevibacter* spp., *Methanomassiliicoccus* spp., and the genus UBA71, greatly expand the number of methanogen species within the human microbiome. Considered the most dominant methanogen in healthy individuals, *M. smithii* can account for up to 10% of the total GIT microbiome and has been detected in up to 96% of subjects (Dridi et al., 2009; Dridi, Raoult, et al., 2011; Eckburg et al., 2005). Taxonomic classification of the MAGs reflects this, with 90% of all recovered genomes classified as *M. smithii*. Interestingly though, 15% of the MAGs were classified as *M. smithii*_A, an uncharacterised subgroup of *M. smithii*. Comparatively, *Methanosphaera* is considered the second most dominant methanogen and is detected in up to 30% of individuals (Dridi, Henry, et al., 2012). However, my recovery of human methanogen MAGs suggests *Methanomassiliicoccales*-associated lineages represent the second most dominant methanogen with these cohorts. Here, eight HQ MAGs were assigned to *Methanomassiliicoccus*, 13 to *Methanomethylophilus*, and 21 classified to the uncharacterised genus UBA71. Comparatively, only four HQ MAGs were assigned to lineages of *Methanosphaera*. Interestingly though, no recovered MAGs were taxonomically classified to *Methanomassiliicoccus luminyensis*, the type strain and first human isolate of *Methanomassiliicoccus* (Dridi, Fardeau, et al., 2012).

Cultured methanogen isolates from human samples have historically been dominated by representatively isolates from North American subjects, typically from the USA. Indeed, the cultured type strains of *M. smithii* (PS) and *M. stadtmanae* (DSMZ3091) were both recovered from the faecal samples of US residents (Miller & Wolin, 1985; Miller et al., 1982). Isolation of representatives from the *Methanomassiliicoccales* has predominantly focused on French subjects (Borrel, Harris, et al., 2013; Borrel et al., 2012; Dridi, Fardeau, et al., 2012). Recently, an additional strain of *M. smithii* (KB11) was also recovered from a Korean faecal sample (Kim & Jeong, 2018). As a part of Chapter 2, I increased the geographical diversity of human *M. smithii* isolates to include two strains recovered from an Australian population. Additionally, I successfully isolated the second only cultured representative of *M. stadtmanae,* also from an Australian individual (Hoedt et al., 2018). Here, the MAGs recovered from publicly available human metagenomes substrantially expand this geographical distribution of human-associated methanogens to included 22 countries from six

continents. This expansion of genomes included MAGs recovered from several diseases and disorders, including those not typically associated with GIT complications. Previous work by Blais Lecours et al. (2014) showed the methanogen population of individuals with CD to shift away from the dominant *M. smithii* and towards *M. stadtmanae*. Additionally, both UC and CD patients showed a reduction in methanogen populations by *McrA* gene count, though the RFLP profiles of *McrA* clones suggested a higher diversity of *M. smithii* and other methanogen species in those with UC (Scanlan et al., 2008). Although no *Methanosphaera* MAGs were recovered from IBD samples, MAGs successfully recovered from CD samples were only assigned to *M. smithii*_A, whereas MAGs assigned to both *M. smithii*_A and *M. smithii* were recovered from patients with UC. This enrichment of certain *M. smithii* lineages in CD could be driven by the changing gut ecology associated with disease progression and the availability of certain metabolites or trace metals required for growth. Anaemia is a well-documented complication of patients with IBD, with up 90% of specific populations affected (Peyrin-Biroulet et al., 2015; Stein et al., 2010). Consequently, the faecal microbiota and biofilms from patients with IBD display a significantly increased concentration of intracellular iron (Motta et al., 2018). Furthermore, depletion of available luminal iron significantly reduces the abundance iron-depend microbes and thus alter the total microbial composition (Werner et al., 2011). Indeed, the genome of CD-associated adherent-invasive *E. coli* strain LF82 contains many virulence factors including genes involved in iron acquisition (Miquel et al., 2010). Similarly, B2 *E. coli* strains isolated from patients with IBD show characteristic traits of extraintestinal pathogenic *E. coli*, including increased iron acquisition system proteins (Petersen et al., 2009). Along these lines, I showed a significant differentiation in metal uptake and transport-associated genes in the different phylogroups of *M. smithii*, with significant enrichment of iron complex transport genes (*ABC.FEV.P, ABC.FEV.S, ABC.FEV.A*) in *M. smithii*_A, and thus possess a greater efficiency of iron sequestration in a CD environment. Comparatively, the enrichment of molybdate transport system genes (*modA-C*) was shown in *M. smithii*, which was recovered at a higher rate in T2D. This differential enrichment of iron and molybdate transport systems may allow for the differential abundance and thus subsequent recovery of MAGs associated with the *M. smithii* phylogroups from CD and T2D patients.

An additional enrichment of genes was also observed between the *M. smithii* phylogroups. *M. smithii* contained a protein annotated as acetoin utilization protein AcuB, which is not well described in archaea but has been characterised in the growth of *B. subtilis* on acetoin and butanediol (Grundy, Waters et al. 1993). Acetate assimilation has been suggested as the primary carbon source of *M. smithii* according to Samuel, Hansen et al. (2007), but the presence of AcuB in *M. smithii* may suggest they are also capable of utilising acetoin or butanediol as a primary carbon source. Interestingly, acetoin has been shown to decrease in acute dextran sodium sulfate-induced mice, as well as chronic

piroxicam-accelerated colitis in interleukin-10$^{-/-}$ mice models (Shimshoni et al., 2020), suggesting this compound may be associated with colitis related diseases. Although scant information is available regarding this observation in humans, the reduction in acetoin may limit carbon sources of specific phylogroups of *M. smithii*. I also showed differential enrichment of genes involved in membrane glycosylation with *M. smithii* enriched for dolichol-phosphate mannosyltransferase DPM1 and *M. smithii*_A enriched for UDP-glucose/galactose:(glucosyl) LPS alpha-1,2-glucosyl/galactosyltransferase, involved in outer membrane lipopolysaccharide glycosylation in *E. coli* (Leipold et al., 2007). Similarly, GH99 is differentially enriched between phylogroups and likely shares endo-α-1,2-mannanase activity involved in the cleavage of glycosylated N-linked glycoproteins (Hakki et al., 2015). A similar strain variation of glycosylation and adhesin-like proteins for *M. smithii* was previously described by Hansen et al. (2011). Phosphonopyruvate decarboxylase, phosphoenolpyruvate phosphomutase, and an alcohol dehydrogenase with phosphonoacetaldehyde reductase homology were also differentially encoded between individual *M. smithii* subgroups. This cluster of genes has been characterised in the biosynthesis of phosphonate compounds such as the antibiotic dehydrophos produced by *Streptomyces luridus* and fosfomycin from *Streptomyces wedmorensis* (Nakashita et al., 1997; Shao et al., 2008; Woodyer et al., 2006). This may suggest specific subgroups of *M. smithii* could utilise this cluster of genes to modulate bacterial species in their surrounding environment. Interestingly, phylogenetic analysis of the genomes produced by Hansen et al. (2011) showed that *M. smithii* TS94A-C, TS95A-D, TS96A-B, TS147A-C, and TS146E were cultured representatives of *M. smithii*_A. As such, the strain variation of *M. smithii* observed by Hansen et al. (2011) represents differences between the unrecognised subgroups of *M. smithii.* Together, the differential genetic enrichment and my phylogenetic assessment warrant the identification of *M. smithii*_A and *M. smithii* as two unique phylogroups of human-associated methanogens with unique genetic adaptations. Importantly, the differentiation between these two phylogroups and their genetic potential is essential in understanding disease-associated shifts in *M. smithii* that may be associated with a specific phylogroup.

MAGs assigned to novel human-associated lineages were also recovered as part of these analyses and provide insight into the novel metabolic potential of human-associated methanogens. *Methanobacterium* was recently recovered from the human GIT by amplicon-based sequencing, though these sequences were only recovered from biopsy samples suggesting this methanogen is mucosal-associated (Koskinen et al., 2017). Here, an HQ MAG (ZellerG_2014_CCIS46047672ST-4-0_bin.2) was assigned to *Methanobacterium*, representing the first available genome of this lineage derived from a human sample. Interestingly, *Methanobacterium formicicum* and *Methanobacterium bryantii* have been shown to utilise secondary alcohols, 2-propanol and 2-butanol, as hydrogen donors

for methanogenesis (Zellner & Winter, 1987). Choline trimethylamine-lyase activating enzymes (*cutC*), catalysing the conversion of choline to TMA and acetaldehyde, were are also encoded (Craciun & Balskus, 2012). As summarised by Kurth et al. (2020), both choline and TMA are utilised by several different lineages of methanogens for methanogenesis and, thus, may similarly be utilised by human-associated *Methanobacterium*. Additionally, a *Methanosphaera* MAG (ZellerG_2014_CCIS33816588ST-4-0_bin.32) was taxonomically classified as *Methanosphaera* sp. RUG761, a close phylogenetic lineage to *Methanosphaera* sp. WGK6 isolated from a Western grey kangaroo (Hoedt et al., 2016). Like *Methanosphaera* sp. WGK6, ZellerG_2014_CCIS33816588ST-4-0_bin.32 also contained homologs of the alcohol and aldehyde dehydrogenase genes allowing for the utilisation of ethanol as a primary substrate for methanogenesis. Together these *Methanosphaera* and *Methanobacterium* MAGs expand our understanding of human-associated methanogens and show a greater potential diversity of substrate utilisation.

Enabled by the recovery of MAGs from publicly available metagenomes as a part of this study and by Pasolli et al. (2019), I have further characterised the human methanogen community and shown novel lineages of human-associated *Methanosphaera*, *Methanobrevibacter, Methanobacterium*, and multiple lineages of *Methanomassiliicoccales*. Additionally, I described a phylogenetic separation of *M. smithii*, showing a smaller uncharacterised phylogroup: *M. smithii*_A. Here I show differentiation in metal update, carbon assimilation, and glycosylation-associated genes between the *M. smithii* phylogroups which, along with their phylogenetic separation, suggest *M. smithii* is in fact two genetically distinct lineages. These phylogroups were also shown to be differentially recovered from diseased samples, with CD samples only producing *M. smithii*_A MAGs, whereas UC samples produced *M. smithii* and *M. smithii*_A MAGs. Collectively, the distinction of these phylogroups and their associated genetic adaptations is integral in understanding the shifts observed for *M. smithii* in IBD, and potentially other diseases and disorders.

3.5 Published research article on work carried out in Chapter 3

Concurrent with my analysis, additional human-derived methanogen MAGs were published by Almeida et al. (2019) and Nayfach et al. (2019). Subsequently, a research paper focusing on the recovery and comparative analysis of the human methanogen MAGs recovered by these studies and Pasolli et al. (2019) was released. As of writing this thesis, the paper by Chibani et al. (2020) is currently in prepublication review and is available as a preprint article. The publication by Chibani et al. (2020) provides an analysis of the human methanogen-associated virome, as well as an assessment of the phylogenetic separation and horizontally transferred genes of the human methanogens compared to those from non-human hosts.

Like my analyses, they show the split in the *M. smithii* and *M. smithii*_A phylogroups, showing a similar enrichment of ABC transporter permease components in *M. smithii* and membrane/cell-wall associated proteins in *M. smithii*_A. However, my analyses expand on these findings, showing the enrichment of additional iron transport-associated protein in *M. smithii*_A, as well as the presence of differential carbon assimilation through the presence of acetoin utilisation protein AcuB in *M. smithii.* I also further characterise the differentially expressed membrane/cell-wall associated proteins using both KO and cazymes annotation. Similarly, I also characterise the enrichment of genes involved in the biosynthesis of phosphonate compounds in *M. smithii*. Although they also characterise MAGs of novel lineages, my work also expands on the metabolic implications of these novel archaeal species.

My work also provided a unique analysis of subgroups within *M. smithii* and *M. smithii*_A, and their associated enrichment of genes. I also show the recovery of *M. smithii*_A MAGs for CD samples, compared with both *M. smithii* phylogroups being recovered from UC samples, and suggest this is a result of the phylogroups being able to differentially respond to and tolerate the ecological changes associated with IBD.

As such, my work will be prepared for publication with a focus on the identification of the *M. smithii* subgroups and their respective genotypes, with an additional focus on the selective enrichment of *M. smithii* in different diseases.

Chapter 4: Isolation and characterisation of novel methanogenic archaea from Australian marsupials

4.1 Introduction

In addition to the growing interest in methane and methanogens in human health and disease, methane is deemed a potent greenhouse gas, estimated to have 28 times the global warming potential of carbon dioxide (Grossi et al., 2018). Animal agriculture is believed to be the largest source of anthropogenic methane emissions (95-109Tg CH_4/year), with ruminant livestock responsible for at least 80% (87-97Tg CH_4/year) of these emissions via feed digestion and associated microbial fermentation (Dangal et al., 2017; Saunois et al., 2016; Wolf et al., 2017). Reducing methane emissions from these food production systems is now imperative from a social, environmental, and economic context. Given the relatively short half-life of atmospheric methane (NCBI, 2021; USEPA, 2012), any reduction in methane emissions is deemed to more rapidly translate into reductions in global warming potential. However, this increasing urgency to reduce livestock methane emissions is matched by the rising global demand in animal products for food, particularly in those regions of the world experiencing improved socioeconomic conditions (Fanzo et al., 2020; Foley et al., 2011; Kc et al., 2018).

While previous research tackling the challenge of livestock methane emissions has overwhelmingly focused on domesticated ruminants and their production systems, methanogens reside within the gastrointestinal tracts of many herbivorous mammals, including humans. Typically, herbivorous diets consisting of foliage or lignin-rich plant products favour microbial communities with increased methanogen diversity and abundance, due to the increased availability of substrates through bacterial hydrolysis and fermentation of plant polysaccharides (St-Pierre & Wright, 2013). However, the limited research with some of Australia's native herbivores suggests many of these animals are "low-methane" emitters. For instance, kangaroos and wallabies (members of the Macropodidae family) are foregut fermenters and eruct less methane (when corrected for digestible energy intake) compared to ruminant livestock when reared on the same diet (Von Engelhardt et al., 1978). Recent work by Vendl et al. (2015) similarly showed two kangaroo species, *Macropus fuliginosus* and *Macropus rufus*, to produce a lower ratio of CH_4/CO_2 compared to the ruminant species examined. However, they do note that the methane yield per unit of dry matter intake, gross energy intake, and intake of digestible neutral detergent fibre was higher than expected, representing intermediate methane yield between low methane-emitting hindgut fermenters and high methane-emitting ruminants (Vendl et al., 2015). Both culture-independent and -dependant methods show that the methanogens present in the ruminant and macropodid foregut are phylogenetically linked, with lineages of *Methanobrevibacter* (e.g. sp. WBY1 from the tammar wallaby, *Macropus eugenii*) and *Methanosphaera* (e.g. sp. WGK6 from the Western grey kangaroo, *Macropus fuliginosus*) recovered in culture, as well as the detection of

Thermoplasmatales- (*Methanomassiliicoccales*) affiliated lineages in 16S rRNA gene amplicon libraries (Evans, 2011; Hoedt et al., 2016). Interestingly though, methanogen communities of the macropodid foregut are much smaller in both relative and absolute abundances compared to ruminants, and in some cases undetectable (Evans et al., 2009; Klieve et al., 2012; Ouwerkerk et al., 2009). Furthermore, the phenotypic and genomic characterisation of *Methanosphaera* spp. shows unique attributes that confirms there are differences between the lineages of methanogens present in ruminant and monogastric hosts (Hoedt et al., 2016), which includes large differences in genome size and content (Hoedt et al., 2018).

The Macropodids represent only one of the extant branches of the Diprotodonts, which captures the diversity of native Australian herbivores (Beck et al., 2020; Dodt et al., 2017; Meredith et al., 2009). For instance, koala species are recognised for their highly selective diet of fresh leaves selected from only a few of the many Australian *Eucalyptus* spp. native to the continent, which undergoes digestion and nutrient release via microbial activity in the hindgut. Wombat species are also hindgut digesters, but their nutritional ecology is broad and includes native grasses, shrubs, and small plants (Casey et al., 2021). Prior to my PhD studies, members of my supervisory committee initiated studies via an Australian Research Council Discovery Project to use metagenomics and culture-based approaches to characterise the gut microbiota of native Australian herbivores. As part of this research, the comparison of the gut (stool) metagenome from a koala (*Phascolarctos cinereus*) and Southern hairy-nosed wombat (*Lasiorhinus latifrons*) showed the wombat possessed a greater relative abundance of reads assigned to the Domain Archaea (2.14%) compared to koalas (0.11%), with *Methanocorpusculum* identified as the dominant methanogen in the wombat samples (Shiffman et al., 2017). The lower abundance in koalas is speculated to be in part due to the presence of plant secondary metabolites derived from the *Eucalyptus* diet, though this remains to be tested (Cieslak et al., 2013). A survey of ~100 stool samples as part of this project confirmed that the koala samples typically produced less methane in culture compared to samples from mahogany gliders (*Petaurus gracilis*) and wombats (Soo, Hoedt, et al., unpublished data). Furthermore, the MGS data produced from a subset of these samples supported the recovery of archaeal MAGs by Dr Rochelle Soo and surprisingly, these MAGs did not represent the "canonical" gut methanogen lineages, but rather, were affiliated with the family *Methanocorpusculaceae.*

The *Methanocorpusculaceae* have long been recognised as "environmental" methanogens, principally isolated from soils and hydrocarbon-rich bogs (Zellner et al., 1989; Zhao et al., 1989). There have also been sporadic reports of the recovery of 16S rRNA gene amplicons affiliated with the *Methanocorpusculaceae* from digesta/stool samples of domesticated animals including chickens, horses, and cattle (Doster et al., 2018; Fernandes et al., 2014; Hou et al., 2016), as well as the land

iguana (*Conolophus subscristatus* and *C. pallidus*)(Hong et al., 2011), captive and wild ptarmigans (*Lagopus muta*)(Salgado-Flores et al., 2019), and white and black rhinoceros (*Ceratotherium simum*, *Diceros bicornis*)(Gibson et al., 2019; Luo et al., 2013). While it is reasonable to suggest these amplicons may have arisen from exogenous sources (e.g. via soil ingestion and/or soil contamination of faecal samples), intriguingly, *Methanocorpusculum*-affiliated lineages have also recently been reported in the 16S rRNA gene amplicon profiles of faecal samples from baleen (*Balaenoptera musculus, B. physalus, B. borealis*), toothed (*Physeter macrocephalus*), and sperm whales (*Physeter catodon*)(Glaeser et al., 2021; Li et al., 2019).

Collectively, the findings outlined above raise the spectre that the biodiversity of gut methanogenic archaea may need to be expanded to include members of the *Methanocorpusculaceae.* However, without the recovery of isolate(s) representing these lineages, these molecular-based observations lack sufficient biological foundations. The resources available to me via the Discovery Project offered me a unique opportunity to not only overcome this knowledge gap, but also overcome the constraints on my Ph.D. research instigated by the SARS-COV-2 pandemic. In this Chapter, I present my findings that validate *Methanocorpusculum* spp. are autochthonous members of the gut microbiota in native Australian herbivores and other vertebrates. To that end, I have produced axenic isolates of this archaeal lineage from stool samples of the common wombat (*Vombatus ursinus*) and mahogany glider (*Petaurus gracilis*) and produced MAGs from the metagenomic datasets of our archived samples from native Australian herbivores. Using these isolates and recovered MAGs, I show the host-derived lineages of *Methanocorpusculum* are different from those of the environmental isolates. Another outcome of my culture-based studies for this Chapter was the recovery of novel *Methanobrevibacter* and *Methanomethylophilaceae* spp. from stool samples of Eastern grey kangaroo (*Macropus giganteus*) and mahogany glider, respectively, which are also briefly described here. These isolates and the subsequent comparative genomics expand on the functional and bioinformatic findings of chapter 2 and 3, and further characterise lineage-specific genetic adaptations for host-associated methanogenic archaea.

4.2 Materials and Methods

4.2.1 Marsupial sample collection and storage

Marsupial faecal samples were collected from sanctuaries and zoos in South-East Queensland (Lone Pine Koala Sanctuary, Brisbane and David Fleay Wildlife Park, Burleigh Heads) and North Queensland (Wildlife Habitat, Port Douglas and Cairns Tropical Zoo) by two individuals (DS Teakle and Amy Shima). Faecal material was collected from 23 marsupial species, including greater glider (Petauroides Volans, n=4), mahogany glider (Petaurus gracilis, n=9), squirrel glider (Petaurus norfolcensis, n=4), yellow-bellied glider (Petaurus australis, n=1), Eastern grey kangaroo (*Macropus giganteus*, n=13) , red kangaroo (*Macropus rufus*, n=12), koala (*Phascolarctos cinereus*, n=125), Lumholtz's tree-kangaroo (*Dendrolagus lumholtzi*, n=7), red-legged pademelon (*Thylogale stigmatica*, n=4), common brushtail possum (*Trichosurus vulpecula*, n=10), common ringtail possum (*Pseudocheirus peregrinus*, n=4), green ringtail possum (*Pseudochirops archeri*, n=5), Herbert River ringtail possum (*Pseudochirulus herbertensis*, n=2), mountain brushtail possum (*Trichosurus cunninghami*, n=1), short-eared possum (*Trichosurus caninus*, n=3), striped possum (*Dactylopsila trivirgata*, n=2), agile wallaby (*Macropus agilis*, n=3), northern nail-tail wallaby (*Onychogalea unguifera*, n=3), parma wallaby (*Macropus parma*, n=3), red-necked wallaby (*Macropus rufogriseus*, n=3), swamp wallaby (*Wallabia bicolor*, n=2), common wombat (*Vombatus ursinus*, n=6), and Southern hairy-nosed wombat (*Lasiorhinus latifrons*, n=8). Ethical permission for the collection of all samples was granted by the Animal Welfare Unit, the University of Queensland, Brisbane, Australia under ANRFA/SCMB/099/14. All samples were stored at –80°C in Eppendorf tubes within 24h and until processing.

4.2.2 Faecal sample DNA extraction, amplicon sequencing, and metagenome sequencing (MGS)

These procedures were performed by Dr Rochelle Soo (The University of Queensland, Australian Centre for Ecogenomics). A subset of collected marsupial faecal samples were chosen for DNA extraction, including: Herbert River ringtail possum (n=2), striped possum (n=2), Lumholtz's tree-kangaroo (n=7), Southern hairy-nosed wombat (n=7), agile wallaby (n=3), Eastern grey kangaroo (n=8), red-necked wallaby (n=2), parma wallaby (n=3), red kangaroo (n=3), northern nail-tail wallaby (n=3), greater glider (n=4), mahogany glider (n=9), squirrel glider (n=4), yellow-bellied glider (n=1), koala (n=10), common ringtail possum (n=4), green ringtail possum (n=5), red-legged pademelon (n=4), common brushtail possum (n=9), mountain brushtail possum (n=1), short-eared possum (n=3), common wombat (n=4), and swamp wallaby (n=2).

The faecal DNA extraction methods were described by Shiffman et al. (2017). In brief, ~50 mg of each faecal sample was combined with 0.7 mm garnet beads and suspended in 750 µL Tissue Lysis Buffer. Samples were then homogenised at 2000 rpm for 5 min using a MoBio Powerlyzer. Samples were then centrifuged at 10,000 x *g* and the supernatant (~300 µL) was collected, and the DNA purified using the Maxwell 16 Tissue DNA Purification Kit and Maxwell 16 Research Instrument (Promega, Madison, WI, USA), as per the manufacturer's instructions. Qubit fluorometer with Quant-it dsDNA BR assays (Invitrogen, Thermo Fischer Scientific) was used to quantify the extracted genomic DNA, with each sample then normalised to 5 ng/mL with sterile water. The V6-V8 hypervariable region of the 16S rRNA genes were amplified by PCR in 50 µL volumes containing 25 ng of DNA, 5 µL of 10x buffer, 1.5 µL of Bovine Serum Albumin (Roche diagnostic, Australia), 0.2 µL of 1 U Fisher Taq DNA polymerase (Thermo Fisher Scientific Inc. USA), 1 µL of dNTP mix (each at a concentration of 10mM), 4 µL of 25 mM MgCl2, 1 µL of each 10 mM of 926F and 1392R primers (Engelbrektson et al., 2010) ligated to Illumina adapter sequences. Each reaction was performed using the following cycling conditions: 95°C for 3 min, followed by 30 cycles of 95°C for 30 s, 55°C for 30 s, 74°C for 30 s, and a final extension at 74°C for 10 min. AMPure XP beads (Beckman Coulter, Rea, CA, USA) were used to purify the resulting amplicons, as per the manufacturer's instructions. Each sample was then indexed with unique 8 bp barcodes using the Illumina Nextera XT V2 Index Kit Set A-D (Illumina FC-131-1002; Illumina, San Diego, CA, USA) under standard PCR conditions. Equimolar indexed amplicons were pooled and sequenced at the Australian Centre for Ecogenomics, using the Illumina MiSeq platform with the version 3 reagent kit for 300 cycles, according to the manufacturer's instructions. The raw data was demultiplexed and processed as per Shiffman et al. (2017).

For the MGS sequencing, aliquots of the extracted DNA were subjected to double size selection for Illumina library preparation. First, 60 µL of AMPure XP beads (Beckman Coulter, Rea, CA, USA) was mixed with 100 µL of the DNA extract, vortexed, and held at room temperature for 5 min. The sample tubes were then placed on a magnetic stand for ~5 min, and once the solution was clear, the supernatant containing the desired DNA fragments was transferred to another sterile tube and the beads discarded. This process was repeated with 10 µL of AMPure XP beads (Beckman Coulter, Rea, CA, USA) and then the sample tube was placed on a magnetic as above, after which the supernatant was discarded. While remaining on the magnetic stand, the beads were washed by two rounds of exposure to 200 µL of 80% (v/v) ethanol for 30 sec, with the ethanol removed at each step via pipette. The beads were then air-dried for ~15 min on the magnetic stand, and 25 µL of nuclease-free water added, vortexed, and held at room temperature for 2 min. The mixtures were then placed on the magnetic stand for ~1 min (or until the solution was clear) and the liquid containing the eluted DNA

was harvested via pipette and transferred to a new sterile tube. The DNA libraries for each sample were constructed using the Illumina Nextera XT DNA library preparation kit (Illumina, San Diego, CA, USA) and ~3 nM of each library was then sequenced using the Illumina NextSeq 500 platform with 2 x 150-bp paired-end chemistry, using standard protocols at the Australian Centre for Ecogenomics.

4.2.3 Assessment of Archaea prevalence and diversity in MGS datasets

The GraftM package (v0.12.2) was used with the forward reads of each MGS dataset to recover those derived from archaeal 16S rRNA genes as identified against 05.2013_08_greengenes_97_otus.gpkg (Boyd et al., 2018). The entire MGS dataset was also examined for archaeal-derived reads and classified using GTDB-Tk classify_wf (v1.0.2) with release r89 (Chaumeil et al., 2019). CoverM (v.0.4.0)(https://github.com/wwood/CoverM) was then used to determine the coverage of reference genomes in the MGS, and heat maps were generated using GraphPad Prism 9. A phylogeny was constructed using the von Willebrand factor (vWF) of respective species available from the NCBI nucleotide database and used to display the detected archaeal species relative to the marsupial hosts. The vWF genes were aligned using MUSCLE in MEGA-X (Kumar et al., 2018), and phylogeny was inferred using Maximum likelihood evaluated with 1000 bootstrap replications.

4.2.4 Recovery of archaeal MAGs from marsupial MGS datasets

The initial set of methanogen MAGs from marsupial gut MGS datasets was produced by Dr. Rochelle Soo as part of the ARC Discovery Project. First, SeqPurge with default settings (v.2018_11)(Sturm et al., 2016) was used for adaptor trimming of the raw shotgun reads. Metaspades v3.13.0 (Nurk et al., 2017) was used for the of contiguous sequences, with auto PHRED offset and k-mer assembly lengths of 21, 33, and 55. BamM v1.7.3 (https://github.com/Ecogenomics/BamM) was used to map the paired-end reads of samples back to respective sample types (kangaroo, koala, possum, or wombat) and used to produce contig coverage. UniteM v0.0.16 (https://github.com/dparks1134/UniteM) was used to recover MAGs from each sample, and CheckM v1.0.12 (Parks et al., 2015) was used to assess the quality of the assembled bins by their estimated completeness and contamination scores. Those MAGs with scores ≥ 50% completeness and ≤ 10% contamination were retained and taxonomically assigned using GTDB-Tk v1.0.2 (Chaumeil et al., 2019). The estimated mean coverage of contigs for each MAG was determined using CoverM v.0.4.0 (https://github.com/wwood/CoverM) in contig mode.

4.2.5 Faecal sample culture for assessment of methane positivity

These procedures were performed by Dr. Emily Hoedt as part of the ARC Discovery Project. A subset of the marsupial faecal samples was chosen for culturing, including koala (Northern n=109, Southern n=2), Southern hairy-nosed wombat (n=5), common wombat (n=3), common brushtail possum (common n=3, golden n=1), mountain brushtail possum (n=1), mahogany glider (n=1), red kangaroo (n=12), Eastern grey kangaroo (n=10), swamp wallaby (n=2), and red-necked wallaby (n=3). Subsamples of the frozen marsupial faecal samples were aseptically transferred to 10 mL of sterile and anaerobically prepared BRN-RF30 broth medium (Hoedt et al., 2016; Joblin et al., 1990) dispensed into Hungate culture tubes and incubated at 37°C for 24 h. Headspace gas (~2 mL) was retrieved from each tube using a sterile needle and syringe and subjected to gas chromatography analysis, as described by Gagen et al. (2014) using a Shimadzu GC-2014 (Shimadzu, Kyoto, Japan) fitted with a flame ionisation detector for CO_2, H_2, and CH_4.

4.2.6 Methanogen enrichment and isolation from marsupial faecal samples

The wombat, mahogany glider and eastern grey kangaroo samples that produced the highest methane for the given species (Section 4.2.5) were chosen for methanogen isolation. In the first round of culturing, a 200 µL subsample of the individual wombat faecal slurry (CW153; Common Wombat sample no. 153) was used to inoculate 10 mL volumes of anaerobic BRN-RF10 medium prepared in Balch tubes, which had also been pressurised to 150 kPa with either H_2:CO_2 (80:20) gas or H_2 gas alone. The cultures receiving H_2 alone were also supplemented with 1% (v/v) combinations of methanol (Sigma-Aldrich; 179337), ethanol (Sigma-Aldrich; E7023), 2-propanol (Sigma-Aldrich; I9516), 1-butanol (Sigma-Aldrich; 360465), 2 M sodium acetate solution or 2 M trimethylamine (TMA) solution. Streptomycin (600 µg/mL), ampicillin (200 µg/mL) and erythromycin (100 µg/mL) were added to all these cultures, before incubation at 37°C with rotational agitation at 100 rpm. Once free of bacterial contamination, CW153 cultures were then 10-fold serially diluted and 0.1 mL volumes of the highest dilution which showed growth was transferred to Hungate roll tubes pressurised with the gas mixes described above and containing 4.5 mL of BRN-RF10 medium containing the same carbon sources as the parent enrichment and 0.7% (w/v) agar, which were kept molten at 65 °C. After inoculation, the culture tubes were then placed on top of a container filled with crushed ice and rolled to disperse and solidify the agar on the inner wall of the culture tube. The roll tubes were incubated at 37°C for 4-6 weeks, until individual colonies were visible. Random colonies were aseptically picked from the agar using a sterile glass Pasteur pipette and transferred to freshly prepared BRN-RF10 broth medium containing the gas/substrate combination used to produce the original enrichment.

A second round of methanogen isolation was also performed using the faecal slurries from a mahogany glider and Eastern grey kangaroo, as above. As above, streptomycin and ampicillin were used to suppress bacterial contamination but were changed for erythromycin (100 μg/mL) and vancomycin (50 μg/mL) after 10 subcultures of the mahogany glider enrichments to further suppress bacterial growth. Here, sterile anaerobic BRN-RF10 medium with 1.5% (w/v) agar was prepared, supplemented with respective substrates used for each enrichment and dispensed into sterile disposable Petri dishes within an anaerobic chamber (Coylab, Michigan, USA) filled with an atmosphere of CO_2:H_2:N_2 (15:5:85). A 100 μL aliquot of bacteria-free enrichment cultures was applied to the surface of these agar plates and incubated at 37°C for 4-6 weeks, until single colonies were visible. Random colonies were picked from the agar surface and propagated via their aseptic transfer to broth medium, as described above. The resulting broth cultures from both rounds of enrichment were then evaluated for their purity and taxonomic origin using those methods outlined in Chapter 2.2.1. using archaeal-specific 16S rRNA PCR (86F/1492R)(Wright & Pimm, 2003). PCR amplicons were cleaned using the Wizard SV Gel and PCR Clean-Up System gel extraction protocol, as per the manufacturer's instructions, and sequenced at AGRF (https://www.agrf.org.au/). Cultures with homogenous morphology, no bacterial contamination and a single 16S rRNA amplicon sequence were determined to be axenic. Subsamples of axenic broth cultures for four novel methanogens, hereafter referred to as *Methanocorpusculum* sp. CW153, *Methanocorpusculum* sp. MG, *Methanomethylophilus* sp. MG2 and *Methanobrevibacter gottschalkii* EGK (see Results for more details) were aseptically transferred to a sterile, anaerobically prepared glycerol solution and stored at -80 C, as described by Teh et al. (2021).

4.2.7 Methanogen whole genome sequencing

Genomic DNA was extracted using the consecutive freeze-thaw method described by Hoedt et al. (2018). With the exception of *M. gottschalkii* strain EGK, cell biomass from 10 mL broth cultures of the individual methanogen strains was harvested by centrifugation at 13000 x *g* for 5 min, resuspended in 0.5 ml of RBB+C lysis buffer (500 mM NaCl, 50 mM Tris-HCl, pH 8.0, 50 mM EDTA, and 4% sodium dodecyl sulphate) and heated at 80°C for 10 min. Samples then underwent 15 sets of consecutive freeze-thaws on dry ice for 5 min and 55°C for 3 min. The mixtures were then incubated with 5 μL of 20 mg/mL proteinase K at 55°C for 30 min, and the DNA purified by 3 rounds of phenol:chloroform extraction. Then, 0.1 vol of 3 M sodium acetate (pH 5.2) was added followed by 1.0 vol of isopropanol and placed on ice for 30 min. The DNA was then pelleted at 13000 x *g* for 5 min, carefully rinsed with 0.1 mL of 70% (v/v) ethanol, recentrifuged and allowed to air dry. The dried pellet was then resuspended in 25 μL of 10 mM Tris-HCl (pH ~8.0), with 5 μL of 10 mg/mL RNase A added and incubated at 37°C for 30 min. The quality and quantity of the genomic DNA

samples was first confirmed using a Nanodrop device, and via agarose gel electrophoresis, prior to genome sequencing at the Australian Centre for Ecogenomics. The Nextera DNA Flex Library Preparation Kit (Illumina #20018705) was used according to the manufacturer's instructions and the Mantis Liquid Handler (Formulatrix) was used for library preparation and clean up. Each library was quality assessed using the TapeStation 4200 (Agilent #G2991AA) with Agilent D1000 HS tapes (#5067-5582) and quantified using the Quant-iT™ dsDNA HS Assay Kit (Invitrogen), as per the manufacturer's instructions. Each library was sequenced using the Illumina NextSeq500 platform with NextSeq 500/550 High Output v2 2 x 150 bp paired-end chemistry, and a sequencing depth of one Gbp for each sample.

Multiple attempts to produce high-quality genomic DNA from *M. gottschalkii* strain EGK using the methods described above were unsuccessful. Instead, the cell biomass was harvested from 100 mL BRN-RF10 medium with a headspace of 150 kPa of H_2:CO_2 (80:20) by centrifugation at 15,000 x *g* for 5 min, and provided to the Australian Centre for Ecogenomics for DNA extraction and subsequent sequencing. Here, ~50–200 mg of the cell biomass was transferred to a tube containing 0.1 mm glass beads (BioSpec Products #11079101) and combined with 750 µL of Bead Solution (Qiagen #12855-100-BS), 60 µL of solution C1, and vortexed. The tubes were then heated at 65°C for 10 min. The sample was then subjected to bead beating at 2000 rpm for five minutes on a Powerlyser 24 homogenizer (Mo-Bio #13155) and then centrifuged at 10,000 x *g* for one minute. The resulting lysate was then extracted using the Qiagen DNeasy Powersoil Kit (cat #12888-100), with a final elution volume of 50 µL. The sample library was prepared, and QC performed as above, with the addition of the Epmotion (Eppendorf # 5075000301) automated platform for preparation and clean-up. The library was sequenced using NovaSeq6000 (Illumina) with NovaSeq6000 SP kit v1.5, 2 x 150 bp paired-end chemistry, as per the manufacturer's protocol. Sequenced samples were trimmed using Trimmomatic (v0.32)(Bolger et al., 2014) and assembled using Spades (v3.14.1) with the 'meta' function (Nurk et al., 2017). Estimated completeness and contamination of each genome assembly was determined using CheckM (v1.1.2)(Parks et al., 2015) and taxonomic classification was performed using GTDB-tk (v1.3.0)(Chaumeil et al., 2019). The coverage of each genome was determined using BamM (v1.7.3)(https://github.com/Ecogenomics/BamM) and samtools mpileup command: awk '{ count++ ; SUM += $4 } END { print "Total: " SUM "\t" "Nucleotides: " count "\t" "Average_coverage: " SUM/count }'. Predicted coding sequences were annotated using prokka (v1.14.6)(Seemann, 2014) and the IMG Annotation Pipeline v5.0.23 (https://img.jgi.doe.gov/submit/)(Chen et al., 2019; Huntemann et al., 2015). BlastKOALA was used to assign Kegg Orthologs to the predicted protein sequences on each genome (M. Kanehisa et al.,

2016). The ProgressiveMauve option within the MAUVE software package was used for the multiple genome alignment of *Methanocorpusculum* isolate genomes (Darling et al., 2010).

4.2.8 Microscopy and transmission electron microscopy of marsupial methanogen isolates

Methanogen isolates were cultured in BRN-RF10 medium with 150 kPa of CO_2:H_2 (20:80) headspace gas and respective substrates for methylotrophic strains. For light microscopy, samples of the cultures were heat-fixed on glass slides and stained using standard Gram staining protocols. Gram-stained slides were then imaged using a Nikon Eclipse 50i, under 100 x magnification. Wet mount slides of each culture were visualised using a Zeiss AX10 epifluorescence microscope at 420 nm with a cyan (47 HE) filter set. Transmission electron microscopy of each isolate was conducted by Rick Webb at the University of Queensland Centre for Microscopy and Microanalysis. Cultures of each isolate were pelleted and mixed with low gelling temperature agarose made with uninoculated BRN-RF10 medium. The sample was then immediately frozen using a Leica EMPACT2 high-pressure freezer. Each sample was then freeze-substituted (1% osmium tetroxide, 0.5% uranyl acetate, and 5% water in acetone), as per McDonald and Webb (2011). Samples were brought to room temperature and washed with acetone. Epon resin was used for infiltration and allowed to polymerise for 2 days at 60°C. A Leica Ultracut UC6 ultramicrotome was used to produce ultrathin sections, which were picked up on Formvar coated copper grids. Sections were stained with Reynolds lead citrate for 1 min, 5% uranyl acetate in 50% ethanol for 2 min, and re-stained in Reynolds lead citrate again for 1 min, with a water wash after each subsequent step (Daddow, 1983). Sections were visualised and micrographs were taken using a Hitachi HT7700 transmission electron microscope operated at 80 kV.

4.2.9 Recovery of *Methanocorpusculum* MAGs from publicly available datasets

5,392 metagenome samples from 97 publicly available datasets were downloaded from the NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra) between 18/05/2019 and 15/12/2020 (see Table 6.15 for the full list of datasets). Each metagenome was trimmed and assembled as per Section 3.2.1. Metabat (v2.12.1) was used to produce genome bins with a minimum contig size of 1500 bps. Bin coverage was estimated using BamM and samtools mpileup command: awk '{ count++ ; SUM += $4 } END { print "Total: " SUM "\t" "Nucleotides: " count "\t" "Average_coverage: " SUM/count }'. Bin quality was assessed using the CheckM (v1.0.7)(Parks et al., 2015) lineage_wf command and GTDB-tk (v1.3.0)(Chaumeil et al., 2019) was used to taxonomically assign each archaeal MAG. MAGs of ≥50% completeness and ≤10% contamination were retained for further analyses. MAGs with ≥50% completeness and ≤10% were classified as medium-quality (MQ) and those with ≥90% completeness and ≤5% contamination were classified as high-quality (HQ). Predicted tRNA and

rRNA counts were annotated using Aragorn and Barrnap within prokka (v1.14.6)(Seemann, 2014), with the Archaeal kingdom modifier. Multiple archaeal MAGs produced from a single metagenome were dereplicated using dRep (v2.4.0)(Olm et al., 2017).

4.2.10 Phylogenetic analysis and average nucleotide identity of *Methanocorpusculum* genomes

The MAGs recovered from animal metagenomes, isolate genomes and *Methanocorpusculum* downloaded from NCBI were taxonomically assigned using GTDB-tk (v1.3.0)(Chaumeil, Mussig et al. 2019). Concatenated marker gene files from 122 archaeal marker genes were produced using GTDB-tk (v1.3.0). Phylogeny of the marker gene files was inferred using FastTree (v2.1.10) and visualised by iToL (https://itol.embl.de/). Average nucleotide identity (ANI) was determined using fastani (v1.1) (Jain et al., 2018). Average amino acid identity (AAI) was determined using the Kostas lab online AAI calculator (http://enve-omics.ce.gatech.edu/aai/).

4.2.11 Comparative analysis of *Methanocorpusculum* isolate genomes and MAGs

The HQ *Methanocorpusculum* MAGs and isolate genomes were included in the comparative genomic analyses. The Bacterial Pan Genome Analysis tool (BPGA) was used to cluster gene families with USEARCH, using the default 50% sequence identity cut-off for orthologous clustering (Chaudhari et al., 2016). Pan-genome profile analysis within BPGA was used to identify the core and pan-genome for all *Methanocorpusculum* using 100 replications. Additionally, subset analyses were used to produce pan-genome data for each *Methanocorpusculum* clade or genome group stratified by environment or host origin. Pan-genome functional analysis was used to assign COG and KEGG Orthology to protein sequences from each clade and visualised using GraphPad Prism 9. PCA plots of gene ortholog variance were generated using EnrichM, using the '--orthologs' annotation and enrichment functions (v0.4.15). Statistical analysis of gene distribution across the *Methanocorpusculum* clades was performed using EnrichM (v0.5.9)(https://github.com/geronimp/enrichM). Genomes were annotated using the EnrichM 'annotate' function, with the input '--ko' for KEGG Orthologs and '--cazy' for Cazymes. EnrichM 'enrichment' was used to determine genes differentially enriched in defined *Methanocorpusculum* groups, with statistical analysis performed in EnrichM by Fisher's Exact test and Mann–Whitney U test. Corrected p-values of <0.05 were considered significant. Cazyme annotations were visualised as a heatmap using iToL (https://itol.embl.de/) against a phylogenetic tree of HQ *Methanocorpusculum*, produced as described in 4.2.10.

4.2.12 *Methanocorpusculum* spp. growth kinetics and substrate utilisation

Growth curves for *Methanocorpusculum* sp. MG and CW153 were conducted in BRN-RF10 medium prepared as per Balch et al. (1979), without the addition of sodium acetate and sodium formate, and prepared using N_2 gas. 10 mL aliquots were prepared in Balch tubes and each pressurised with CO_2, H_2, or CO_2/H_2 (20:80) headspace gas at 150 kPa. Cultures with CO_2/H_2 were used as positive controls and CO_2 or H_2 alone was used as the negative control. Substrate utilisation test cultures contained either CO_2 or H_2, along with 1% v/v supplementation of 2 M sodium acetate solution (Sigma-Aldrich; S2289), 2M sodium formate solution (Sigma-Aldrich; 798630), 2M methylamine solution(Sigma-Aldrich; M0505), methanol (Sigma-Aldrich; 179337), ethanol (Sigma-Aldrich; E7023), 1-propanol (Sigma-Aldrich; 402893), 2-propanol (Sigma-Aldrich; I9516), 1-butanol (Sigma-Aldrich; 360465), 2-butanol (Sigma-Aldrich; 19440), iso-butanol (Sigma-Aldrich; 320048), tert-butanol (Sigma-Aldrich; 360538), 1-pentanol (Sigma-Aldrich; 76929), 2-pentanol (Sigma-Aldrich; P8017), cyclopentanol (Sigma-Aldrich; C112208), cyclohexanol (Sigma-Aldrich; 105899), 2,3-butanediol (Sigma-Aldrich; B84904), or glycerol (Chem Supply; GA010). Parent cultures of *Methanocorpusculum* sp. MG and CW153 were grown to mid-exponential phase (0.2 OD_{600}) and 200 µL was aseptically inoculated into each prepared Balch tube. Cultures were inoculated in triplicate and incubated horizontally at 37 °C with 100 rpm rotational agitation. Growth was measured by OD_{600} at two-hourly intervals for ~36 h and then every 24 h there after until ~500 h. Growth curves were visualised using GraphPad Prism 9.

4.3 Results

4.3.1 Gas production from marsupial faecal samples

As shown in Figure 4.1, gas production varied greatly across the faecal samples from the different marsupial species. While CO_2 production was observed across all faecal cultures tested, the percentage of CO_2 was comparatively lower in the koala cultures. Methane was detected at a substantial percentage from the wombat (SHNW - 2.136±0.231%, CW - 3.835±0.336%), kangaroo (RK - 2.349±0.530%, EGK - 1.645±1.021%), wallaby (SW - 1.986±0.883%, RNW - 2.400±1.131%, and mahogany glider (1.042%) cultures, with lower concentrations in the common brushtail possum (0.174±0.270%) and only trace methane concentrations detected in the samples from koala (Northern - 0.033±0.050%, Southern - 0.023±0.008%). Comparatively, most possum cultures instead produced a greater amount of hydrogen and reduced methane, although only a single culture was representative

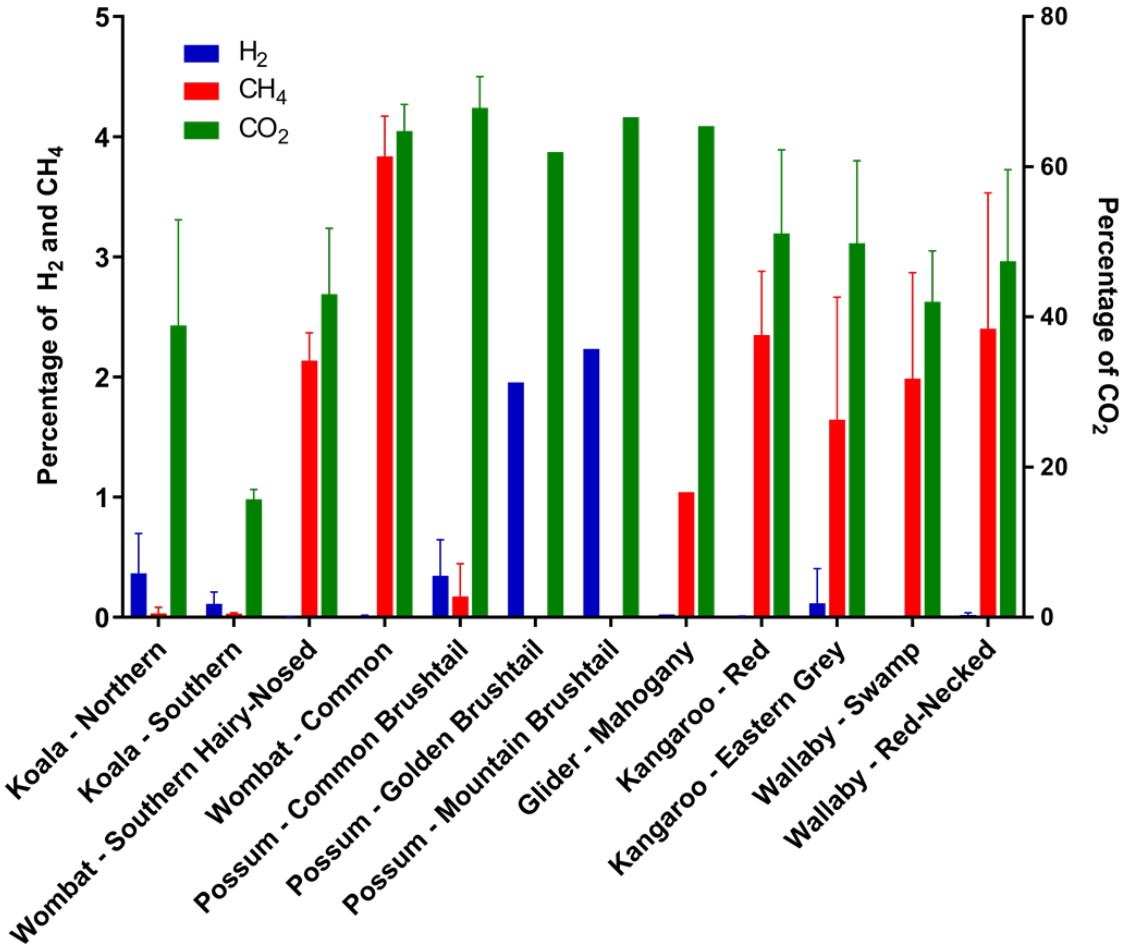


Figure 4.1. Hydrogen, methane, and carbon dioxide gas production by anaerobic cultures inoculated with marsupial faecal samples. Gas production was analysed after 24h incubation using gas chromatography. Values represent percentages of culture headspace gas samples. Kangaroo, wallaby, and wombat faecal cultures are shown to produce a higher concentration of methane compared to possum and koala.

of the mountain and golden brushtail. For the cultures produced from koala faecal samples, the majority (72/111, ~65%) produced only trace amounts of methane, and the amount of hydrogen was also comparatively low. Interestingly, those cultures that produced a greater percentage of methane contained only trace hydrogen, suggesting that at least some of these animals possess a methanogenic consortium.

4.3.2 Methanogen prevalence and diversity in marsupial stool samples

The 16S rRNA gene amplicon profiles suggested three methanogen populations were present in the marsupial faecal samples: *Methanobrevibacter*, *Methanocorpusculum*_1, and *Methanocorpusculum*_2 (Figure 4.2). Only one *Methanobrevibacter* OTU was detected across the samples and showed high sequence homology (>98.81%) to *Methanobrevibacter gottschalkii* HO. Two OTUs of *Methanocorpusculum* with 99.6% sequence similarity were also detected, which to my knowledge has not been previously reported, and showed 99.2% sequence similarly to *Methanocorpusculum labreanum* Z.

Of the marsupial faecal samples, 68% (69/102) contained at least one detectable methanogen OTU (see Table 6.11-12 for complete list). Almost all Eastern grey and red kangaroo samples contained *Methanobrevibacter* (10/11) and *Methanocorpusculum*_2 (10/11). All wallaby species showed a similar high percentage of *Methanobrevibacter* and *Methanocorpusculum*_2, except for the Northern nail-tail wallaby, which contained no detectable *Methanocorpusculum*_2. As divergent members of Macropodidae, Lumholtz's tree kangaroo and red-legged pademelon showed a greater prevalence of *Methanocorpusculum*_2 (5/7, 4/4, respectively) compared to *Methanobrevibacter* (2/7, 2/4, respectively). *Methanocorpusculum*_1 also showed a low prevalence in Lumholtz's tree kangaroo (2/7) and red-legged pademelon (1/4) samples. The common brushtail possum showed *Methanobrevibacter* in 3/9 samples with a low abundance (0.0014± 0.0003%), similar to the short-eared possum which only showed *Methanocorpusculum*_2 in 1/3 faecal samples. Interestingly, the green ringtail, Herbet River ringtail and mountain brushtail contained no detectable methanogens. The mahogany and squirrel gliders showed a high prevalence of *Methanocorpusculum* OTUs, with only a single mahogany glider (1/9) showing a low abundance of *Methanobrevibacter* (0.00194%). Indeed, all squirrel gliders (4/4) and 78% (7/9) of mahogany gliders contained *Methanocorpusculum*_1 and *Methanocorpusculum*_2. Interestingly, mahogany glider samples contained the highest average abundance of *Methanocorpusculum*_1 at 0.7± 1.0%, as well as the highest individual sample abundance of 2.5%. The common and Southern hairy-nosed wombat

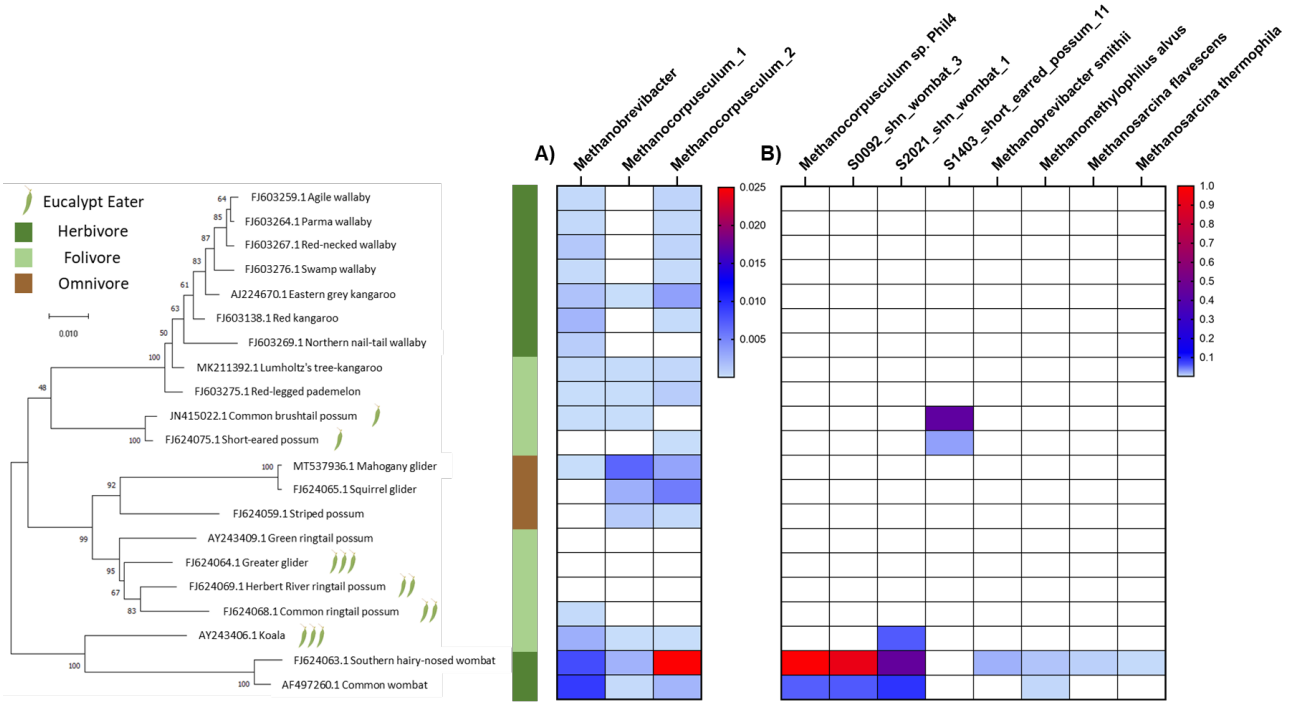


Figure 4.2. Methanogen profiles detected in marsupial species with amplicon and metagenomic sequencing. The phylogenetic tree was built using the von Willebrand factor (vWF) of the respective species available from the NCBI nucleotide database. MEGA-X with MUSCLE were used to align the genes, with Maximum-likelihood and 1000 bootstraps used for phylogeny. CoverM (v0.6.0) was used to detected methanogens based on the sequence coverage of reference genomes. Empty cells indicate no methanogen signal was detected. The host diet composition is displayed as per legend, reworked from Shiffman et al., 2017. **A)** represents methanogens detected by amplicon-based sequencing and **B)** represents methanogens detected by metagenomic sequencing, where genomes from the GTDB-tk database and MAGs recovered from the marsupial metagenomes were used as reference. S0092_shn_wombat_3 was classified to the genus *Methanocorpusculum*, S2021_shn_wombat_3 was classified to the order *Methanomassiliicoccales* and S1403_short_earred_possum_11 was classified to the *Methanomethylophilaceae* genus UBA71.

samples were also enriched for *Methanocorpusculum*_1 and *Methanocorpusculum*_2. All common wombat samples contained both OTUs (4/4), with 6/7 Southern hairy-nosed wombat samples containing *Methanocorpusculum*_2 and 5/7 containing *Methanocorpusculum*_1. Interestingly, the Southern hairy-nosed wombat contained a higher abundance of *Methanocorpusculum*_2 compared to the common wombat (2.49±2.14%, 0.229±0.223%), as well as *Methanocorpusculum*_1 (0.243±0.216%, 0.017±0.017%). *Methanobrevibacter* was also detected in the faecal samples from both wombat species, although at a lower prevalence than the *Methanocorpusculum*. Faecal samples from koalas also contained all three methanogens, and interestingly the average concentration of

Methanobrevibacter was greater than that of the kangaroo and wallaby. This comparatively high abundance of *Methanobrevibacter* was attributed to a single koala sample that contained 2.6%, greater than any other marsupial faecal sample, except for those from the wombat samples. Despite the koala displaying all three detected methanogen species, only 3/10 koala faecal samples were positive for *Methanobrevibacter* and only two samples were positive for *Methanocorpusculum*_1 and *Methanocorpusculum*_2, respectively. As discussed in Section 4.3.1, trace concentrations of methane would suggest that the koala faecal samples contain a low abundance of methanogens that is not accurately captured for all samples.

In contrast, the MGS datasets revealed a greater diversity than the 16S rRNA amplicon profiles, with reads assigned to *Methanocorpusculum* spp., *Methanomassiliicoccales*, *Methanomethylophilaceae* g_UBA71, *M. smithii*, *M. alvus, Methanosarcina flavescens,* and *Methanosarcina thermophila* (see Table 6.13-14 for complete list). All methanogens were detected in the Southern hairy-nosed wombat samples, except for g_UBA71, with *Methanocorpusculum* again being the dominant species (0.987102994, 0.924668943, respectively). Similarly, the common wombat samples also contained no g_UBA71, although *Methanosarcina* or *M. smithii* were also not detected. Apart from the wombats, the koala samples contained only novel *Methanomassiliicoccales*. The only other samples with detectable methanogens were the common brushtail and short-eared possum, which contained the only detected g_UBA71 among the faecal samples. Despite *Methanobrevibacter* and *Methanocorpusculum* spp. being detected using 16S rRNA amplicon-based sequencing, no methanogens were detected in the other possum spp., kangaroo, wallaby or glider samples using MGS.

Of the samples chosen for sequencing, higher methane production (Section 4.3.1) did not necessarily correlate with a greater abundance or diversity of methanogen by amplicon or MGS. On average, the cultures inoculated with the red-necked wallaby and red kangaroo faecal samples produced greater methane (3.052±0.089%, 2.757±0.494%) compared to the common and Southern hairy-nosed wombat (2.037±2.880%, 2.137±0.231%). However, the wombat samples contained a greater abundance of methanogens, specifically *Methanocorpusculum*_1 and *Methanocorpusculum*_2. Similarly, a sample from the common brushtail possum produced substantial methane (0.48611%) but no methanogens were detected by amplicon or MGS. This was similarly shown for multiple koala samples. The converse observation was made for different common brushtail possum and common wombat samples, in which methanogens were detected but culture of the faecal samples produced no methane.

In summary, these results show methanogens are detected in faecal samples of a wide variety of marsupial species. Interestingly though, different species appear to be enriched for different lineages of *Methanobrevibacter* and *Methanocorpusculum*. Indeed, I have shown that the faecal samples of

the Southern hairy-nosed wombat contained the largest abundance of *Methanocorpusculum*_2, where feacal samples of the mahogany glider contained the largest abundance of *Methanocorpusculum*_1. The MGS was able to further expand on the diversity of detectable methanogens for the wombats, with many additional species not captured by 16S rRNA amplicon-based sequencing. However, this could not be said for the majority of other marsupial species which showed no detectable methanogens. Based on these results, coupled with the methane production data shown in Section 4.3.1, I chose to use select samples for the targeted isolation of some of the diverse gut methanogens identified via these analyses.

4.3.3 Enrichment and isolation of methanogenic archaea

None of the cultures inoculated with koala faecal samples showed any discernible growth with any of the substrates tested here and produced no measurable methane in headspace gases. While UV-microscopy showed some samples possessed a sparse amount auto-fluorescent "forms", which might represent methanogenic archaea, based on these collective results I chose to discontinue my efforts with these samples.

Microbial growth coupled with methane positivity was the greatest with four samples: the H_2/sodium acetate substrate combination inoculated with the faecal sample from the common wombat (CW); CO_2/H_2 and TMA/methanol/H_2 enrichments started from the same sample of mahogany glider (MG and MG2, respectively) and the methanol/ethanol/CO_2/H_2 combination inoculated with the sample from an Eastern grey kangaroo (EGK2). Except for the enrichment culture produced using TMA/methanol/H_2, all the others possessed a high density of auto-fluorescent cells.

Once the enrichment cultures were determined to be free of bacterial contamination by bacteria-specific PCR, 16S rRNA gene amplicons were produced from the four enrichment cultures using Archaeal Domain primers (Figure 4.3). The amplicon from the EGK2 enrichment was most closely related to *Methanobrevibacter gottschalkii* sequences produced from strain PG isolated from swine (98.87% identity) and strain HO from horse (98.3% identity)(Miller & Lin, 2002). The amplicon from the MG2 enrichment was assigned to a lineage within the *Methanomassiliicoccales,* and clustered with sequences derived from the "methanogenic archaeon mixed culture" ISO4-G1 (Kelly et al., 2016) and "M. archaeon" DOK (Padmanabha et al., 2013), recovered from sheep and chicken, respectively. Both the MG and CW enrichments produced amplicons that clustered within the Order *Methanomicrobiales,* forming a deep lineage to the other available *Methanocorpusculum* 16S rRNA sequences. The amplicon from the MG enrichment showed only 96.38% sequence similarity to an uncultured archaeon clone HCe_seq85f recovered from equine hindgut (GenBank: MG585205.1) and

the amplicon from the CW enrichment possessed 96.34% identity to uncultured archaeon clone HA-E2, produced from artesian spring and deep sedimentary aquifers (GenBank: AB288242.1).

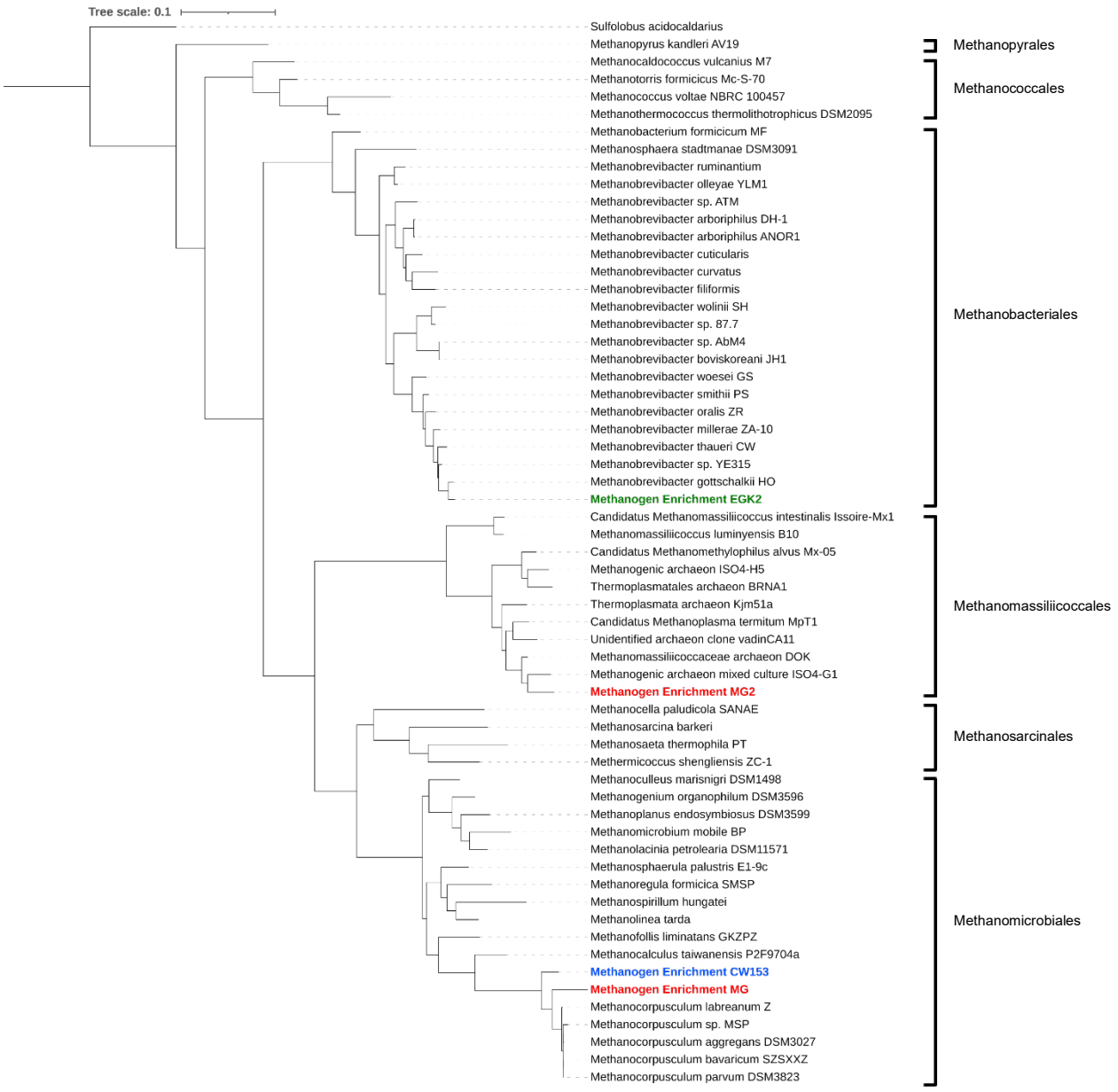


Figure 4.3. Phylogenetic tree showing the preliminary taxonomic classification of marsupial methanogen enrichment cultures. 16S rRNA amplicons were generated from respective isolates using the 86F/1392R archaeal primers. MEGA-X was then used to align the amplicon sequences with reference methanogen 16S rRNA sequences downloaded from the NCBI nucleotide database. Sequences were aligned using MUSCLE, and phylogeny inferred using Maximum-likelihood and 1000 bootstraps. The kangaroo enrichment is displayed in green, mahogany glider in red and wombat in blue. The scale bar represents 10% sequence divergence and *sulfolobus acidocaldarius* was used as the outgroup. The methanogen enrichments produced amplicons that phylogenetically clustered with *Methanocorpusculum*, *Methanomassiliicoccales* and *Methanobrevibacter*.

In summary, I have expanded the diversity of methanogenic archaea cultured from the gut microbiomes of native Australian herbivores, including two novel representatives of the *Methanocorpusculaceae,* one from the common wombat (CW153) and the other from mahogany glider (MG). These appear to be the first cultured representatives of host-associated *Methanocorpusculaceae* and validate the host-microbe association inferred from 16S rRNA gene amplicon surveys presented here and for other vertebrate gut microbiomes (Hong et al., 2011; Luo et al., 2013; Salgado-Flores et al., 2019). The results presented for the remainder of this chapter focus on these novel *Methanocorpusculum* isolates, and my efforts to expand our understanding of this new group of host-associated methanogens, using culture-based, microscopic, and (meta)genomics approaches.

4.3.4 *Methanocorpusculum* sp. CW153 and MG whole genome phylogeny and analysis

Methanocorpusculum sp. MG and CW153 produced genomes containing 2,026,066 and 1,944,273 bps respectively, larger than the ~1.7-1.6 Mbps genome size of the environmental *Methanocorpusculum* genomes (Table 4.1). Each genome contained 59 and 47 contigs, with an average coverage of 199x and 177x, respectively. Both genomes were high-quality, with *Methanocorpusculum* sp. MG estimated to be 98.01% complete with 1.31% contamination and *Methanocorpusculum* sp. CW153 estimated to be 97.69% completeness with 1.96% contamination. Both strains produced no detectable strain heterogeneity, given that the cultures were produced from single colonies. Using GTDB-tk, the *Methanocorpusculum* sp. CW153 draft genome was taxonomically classified as species_*Methanocorpusculum* sp001940805, which is the MAG recovered from the wombat metagenome Phil4 (Shiffman et al., 2017). In contrast, the *Methanocorpusculum* sp. MG draft genome was only classified to genus_*Methanocorpusculum*, further suggesting that it too represents a novel lineage.

While reordering the contigs according to *M. labreanum* Z, I improved the degree of gene synteny among the genomes from the three environmental isolates and showed a great degree of variation in the host-associated genomes, both in comparison to each other, and the environmental isolates (Figure 4.5). Both the MG and CW153 genomes contain unique non-homologous regions, with the MG genome containing a large xenologous region of ~50 Kbps, primarily containing gene sequences with hypothetical annotations. These variations also affected the ANI scores calculated for the different genomes, with a distinct separation of the environmental and host-associated genomes observed (Figure 4.4). Both *Methanocorpusculum* sp. CW153 and MG show a maximum ANI of 78% to any environmental isolate, suggesting that these isolates represent novel species. Interestingly, the

Table 4.1. Preliminary genome details of *Methanocorpusculum* isolates genomes. Genome quality and details were determined using CheckM (v1.1.2). tRNA and rRNA gene counts were annotated using Prokka (v1.14.6). Taxonomic classification was determined using GTDB-tk (v.1.3.0).

Genome	*M. labreanum*	*M. parvum*	*M. bavaricum*	*M.* sp. CW153	*M.* sp. MG
Strain Designation	Z	XII	SZSXXZ	CW153	MG
Completeness	99.54	98.21	98.21	97.69	98.01
Contamination	0	0	0.66	1.96	1.31
Heterogeneity	0	0	0	0	0
Coverage	-	213.00	-	177.10	199.461
Genome Size	1804962	1709133	1702624	1944273	2026066
Contigs	1	47	34	47	59
N50 (contigs)	1804962	74097	161459	164607	87295
GC (%)	50.00659294	51.4284728	51.42874427	53.3461093	51.9950485
Predicted Genes	1816	1732	1748	1962	2087
Isolation source	Tar Pit Lake Sediment	Anaerobic sour whey digester	Muddy sediment of the wastewater pond	*Vombatus ursinus*	*Petaurus gracilis*
Geography	USA	Germany	Germany	Australia	Australia
GTDB Classification	s__Methanocorpusculum labreanum	s__Methanocorpusculum parvum	s__Methanocorpusculum parvum	s__Methanocorpusculum sp001940805	g__Methanocorpusculum;s

CW153 and MG genomes also show only 88% identity to each other, suggesting the isolates represent two distinct lineages of host-associated *Methanocorpusculaceae*. Indeed, using ~95% cut-off for species demarcation (Richter & Rosselló-Móra, 2009), the environmental *M. parvum* and *M. bavaricum* genomes with an ANI score of 98% likely represent different strains of the same species, whereas the ANI scores for the *M. labreanum* genome is no more than 85% ANI with any other strain, suggesting it also represents a unique lineage. As noted by Barco et al. (2020), ANI thresholds for genus delineation are quite variable but given the ANI scores for the two host-associated genomes

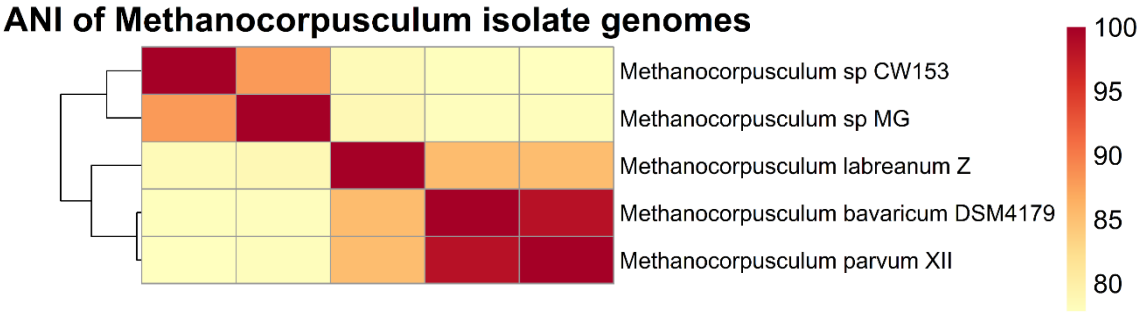


Figure 4.4. Average nucleotide identity (ANI) of *Methanocorpusculum* isolates. Average nucleotide identity (ANI) was determined using fastANI (v1.1)(Jain, Rodriguez-R et al. 2018). The analysis shows a distinct separation between environmental and host-associated *Methanocorpusculum* isolates.

with any environmental isolate is no more than 78% suggests, at a minimum, they are a truly divergent, host-associated clade of the *Methanocorpusculum* genus.

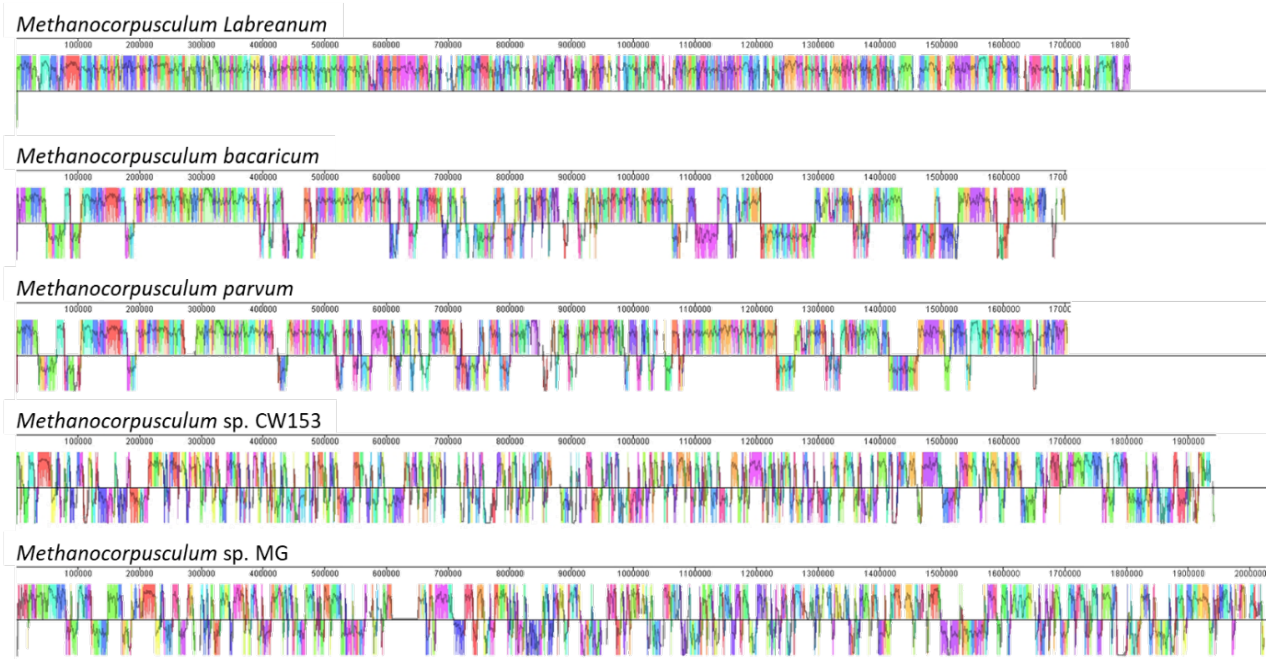


Figure 4.5. Multiple genome alignment of cultured *Methanocorpusculum* genomes and novel isolates *Methanocorpusculum* sp. MG and CW153. Genomes were aligned using ProgressiveMauve in MAUVE (v2.4.0), reordered to the *M. labreanum* Z genome.

Analysis of predicted coding genes

The IMG-based "high-quality" annotations predict that 2104/2162 genes encode proteins for *Methanocorpusculum* sp. MG and 1983/2041 genes encode proteins for *Methanocorpusculum* sp. CW153 (Table 4.2). Both genomes encoded 58 ribosomal genes, with only a single copy of the 5S, 16S, and 23S rRNA genes. Both genomes contained 53 annotated tRNA genes, with tRNAs identified for all standard amino acids except tryptophan. No tRNAs were predicted for pyrrolysine, although a single uncharacterised tRNA sequence was predicted. Functional prediction was assigned to almost 75% of the gene products encoded by both genomes, although less than 50% of the predicted gene products were assigned to KEGG Orthology groups, and only 28% were connected to recognised KEGG pathways. Both genomes shared a similar percentage of genes containing predicted signal peptides (~5%) and transmembrane proteins (~25%), although *Methanocorpusculum* sp. MG contained a larger total number of predicted genes in both categories. In summary, the basic genome characteristics of both genomes are highly similar to each other.

Table 4.2. JGI IMG genome annotation of *Methanocorpusculum* sp. MG and CW153. *Methanocorpusculum* sp. MG and CW153 genomes were annotated using the IMG Annotation Pipeline v5.0.23.

	***Methanocorpusculum* sp. MG**		***Methanocorpusculum* sp. CW153**	
	Number	**% of Total**	**Number**	**% of Total**
DNA, total number of bases	2026067	100.00%	1944274	100.00%
DNA coding number of bases	1804919	89.08%	1719030	88.42%
DNA G+C number of bases	1053455	52.00%	1037195	53.35%
DNA scaffolds	59	100.00%	47	100.00%
Genes total number	2162	100.00%	2041	100.00%
Protein coding genes	2104	97.32%	1983	97.16%
RNA genes	58	2.68%	58	2.84%
rRNA genes	3	0.14%	3	0.15%
5S rRNA	1	0.05%	1	0.05%
16S rRNA	1	0.05%	1	0.05%
23S rRNA	1	0.05%	1	0.05%
tRNA genes	53	2.45%	53	2.60%
Other RNA genes	2	0.09%	2	0.10%
Protein coding genes with function prediction	1560	72.16%	1495	73.25%
without function prediction	544	25.16%	488	23.91%
Protein coding genes with enzymes	534	24.70%	531	26.02%
Protein coding genes connected to KEGG pathways	584	27.01%	582	28.52%
not connected to KEGG pathways	1520	70.31%	1401	68.64%
Protein coding genes connected to KEGG Orthology (KO)	985	45.56%	970	47.53%
not connected to KEGG Orthology (KO)	1119	51.76%	1013	49.63%
Protein coding genes connected to MetaCyc pathways	483	22.34%	480	23.52%
not connected to MetaCyc pathways	1621	74.98%	1503	73.64%
Protein coding genes with COGs	1548	71.60%	1493	73.15%
with Pfam	1504	69.57%	1456	71.34%
with TIGRfam	638	29.51%	607	29.74%
with SMART	417	19.29%	407	19.94%
with SUPERFam	1713	79.23%	1599	78.34%
with CATH FunFam	1316	60.87%	1262	61.83%
in internal clusters	432	19.98%	439	21.51%
in Chromosomal Cassette	2123	98.20%	1994	97.70%
Chromosomal Cassettes	190	-	172	-
Protein coding genes coding signal peptides	115	5.32%	108	5.29%
Protein coding genes coding transmembrane proteins	536	24.79%	512	25.09%
COG clusters	1114	71.96%	1098	73.54%
KOG clusters		0.00%		0.00%
Pfam clusters	1179	78.39%	1146	78.71%
TIGRfam clusters	539	84.48%	520	85.67%

Methanogenesis

The genomes of both *Methanocorpusculum* sp. MG and CW153 contain all necessary genes for methanogenesis by CO_2 and H_2, as well as formate, with both genomes containing the formate transporter *fdhC* and formate dehydrogenase subunits *fdhAB*. *Methanocorpusculum* sp. CW153

contained three copies of *fdhA* and *fdhB*, with two *fdhAB* encoded as consecutive genes and all *fdhAB* components located within a single 22 Kbp region of the genome. Comparatively, *Methanocorpusculum* sp. MG contained three copies of *fdhA* with only two located within 15 Kbps and a single copy of *fdhB*. Although the two closely encoded *fdhA* are both located on contig 30, the other is located near the break of contig 8 and thus likely does not accurately reflect their position with the complete genome. Hydrogenases found in both genomes are similar to those previously reported from studies with *Methanocorpusculum parvum*, except both contain the additional *frhD* and *ehbD* subunits, with MG also containing two copies of the *frhB* subunit (Gilmore et al., 2017). In terms of *frhD*, this is a maturation protease of the F420-dependent hydrogenase and thus may suggest differential methods of maturation (Mand et al., 2018), while the additional subunits may also provide a greater efficiency of transcription. Both genomes contain all five necessary components of *cofCDEHG* for the biosynthesis of coenzyme $F_{420,}$ as expected for hydrogenotrophic methanogens (Thauer, 1998).

The *Methanocorpusculum* sp. CW153 genome encodes all subunits of formylmethanofuran dehydrogenase (*fwd*) except for *fwdH* and *fwdG*. Subunits A-D are encoded in two separate clusters, with all three copies of *fwdE* encoded separately. The third copies of *fwdA-D* are encoded on smaller contigs or close to contig breaks, suggesting that they could form an addition cluster. In the MG genome, one set of *fwdABC* clusters closely with energy-conserving hydrogenases (*Eha*), as has previously been shown in *M. parvum* (Gilmore et al., 2017). This region of CW153 occurs on a break in contig 20 though contig 23 partially encodes for the truncated section, suggesting the two regions may be clustered together. Interestingly, one cluster of *fwd* is grouped with formate dehydrogenase encoding components *fdhAB*, suggesting this may be coregulated for the utilisation of formate in methanogenesis. Similarly, *Methanocorpusculum* sp. CW153 shows three separate clusters of *fwd*, though only one contained the single encoded copy of *fwdD*, along with formate dehydrogenase subunits. Further, *fwdH* and *fwdG* were also not present in *Methanocorpusculum* sp. CW153. Methyl-coenzyme M reductase (*mcr*) and tetrahydromethanopterin S-methyltransferase (*mtr*) subunits were encoded in a single cluster in both genomes, with two copies of *mcrA2* encoded separately in each genome. Both genomes encoded two copies of *mtrA* and no copy of *mtrG*. Heterodisulfide reductase subunits A2/B2/C2 are encoded by MG, in a single cluster, with an additional copy of *hdrA2*. Comparatively, CW153 contains only a single copy of *hdrA2* and a gene with a secondary annotation of *hdrB2*. Neither genome encoded for Coenzyme B biosynthesis, except for MG, which encoded methanogen homocitrate synthase (*aksA*, K10977) as the only component. Similarly, sulfopyruvate decarboxylase (*comDE*, K24393) was the only component encoded for Coenzyme M biosynthesis (CoM), suggesting neither can synthesis Coenzyme B or M by conventional biosynthesis pathways.

Methanocorpusculum sp. MG contains a protein annotated as methylamine-glutamate N-methyltransferase subunit C (MgsC) with predicted involvement in methane metabolism. This gene shares ~70% sequence identity (E=0.0) with rubredoxin protein sequences from *Christensenella* and *Pseudoflavonifractor*. Alpha-hydroxy-acid oxidizing proteins of *Candidatus* Methanomethylophilaceae archaeon RGIG8206 (MBO4552550.1), a MAG recovered from the metagenome of a water buffalo, also show 63% sequence similarity. However, given the absence of *mgsAB* and *mgdABCD*, it is unclear if the MgsC of *Methanocorpusculum* sp. MG plays an active role in methanogenesis. It is also worth noting, neither *Methanocorpusculum* isolate contained homologs of the *adh* shown to allow *M. parvum* to utilise short-chain alcohols in CO_2-dependent methanogenesis (Gilmore et al., 2017).

In summary, the *Methanocorpusculum* genomes examined here share much in common with each other in terms of methanogenesis. Both genomes appear to be restricted to hydrogenotrophic methanogenesis, with no genes encoded for methylotrophic or acetoclastic methanogenesis. However, differences in subunits of formate dehydrogenase and other hydrogenase genes suggests differences in the efficiency of methanogenesis.

Nitrogen, ammonia, and sulphur metabolism

Methanocorpusculum sp. MG contains all necessary genes of the nitrogenase molybdenum-iron protein (*nifDHK*), as well as the additional *nifN* subunit and nitrogen regulatory protein PII 1 and 2, suggesting it can effectively fix nitrogen to ammonia. These genes form a single cluster containing both the *nif* subunits and regulatory proteins. Comparatively, *Methanocorpusculum* sp. CW153 contains only the *nifH* subunit, as well as nitrogen fixation protein *nifB* and *nifU*. Both species also contained a predicted nitrilase gene (Nitrile aminohydrolase, K01501, E3.5.5.1) for the conversion of nitrile compounds to ammonia and carboxylate, which may additionally substitute for the incomplete nitrogenase pathway predicted for *Methanocorpusculum* sp. CW153. Additionally, *Methanocorpusculum* sp. MG and CW153 both contained genes annotated as glutamate dehydrogenase (NAD(P)+) (K00261) and the small chain of glutamate synthase (NADPH)(K00266, *gltD*), as well as two copies of glutamine synthetase (*glnA*), involved in ammonia assimilation. The large subunit of glutamate synthase (NADPH)(K00265, *gltB*) was not present in either genome. Additionally, neither genome contained alanine dehydrogenase for the utilisation of alanine as a nitrogen source. Together this suggests *Methanocorpusculum* sp. MG and CW153 may encode for differential pathways of nitrogen assimilation. In terms of sulphur metabolism, both MG and CW153 contained phosphoadenosine:phosphosulphate reductase (*cysH*, K00390), which is necessary for the production of sulphite from thioredoxin and 3'-Phosphoadenylyl sulphate (*PAPS*)(Berndt et al., 2004).

Both genomes also contain the C subunit of anaerobic sulphite reductase but did not encode for the A/B subunits, though *Methanocorpusculum* sp. CW153 contained two consecutively encoded copies. Together this suggests *Methanocorpusculum* sp. MG encodes more diverse pathways of nitrogen fixation that may translate to greater efficiency compared with *Methanocorpusculum* sp. CW153. This could reflect the gastrointestinal environment of the different animals in which *Methanocorpusculum* sp. MG requires a larger array of pathways to meet the nitrogen requirements.

Amino acid synthesis and uptake

The *Methanocorpusculum* sp. MG genome is predicted to possess a greater number of complete amino acid biosynthesis pathways in comparison to the CW153 genome, with complete pathways for arginine, lysine, leucine/isoleucine/valine predicted. Threonine synthase (*thrC*) is encoded but subunits of ATP:L-homoserine O-phosphotransferase are not, resulting in an incomplete synthesis pathway from serine. Similarly, the beta chain of tryptophan synthase (*trpB*) was present, but the tryptophan biosynthesis pathway was incomplete. Histidine biosynthesis from 5-Phospho-alpha-D-ribose 1-diphosphate (*PRPP*) is also incomplete, as is tyrosine and phenylalanine biosynthesis from chorismate. There was no evidence of genes associated with proline or cysteine biosynthesis, though two copies of 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (*metE*) were encoded, allowing for the biosynthesis of methionine from homoserine. *Methanocorpusculum* sp. CW153 shows a similar propensity of amino acid biosynthesis, except for the absence of dihydroxy-acid dehydratase (*ilvD*) and acetolactate synthase (*ilvB, ilvM, ilvH*) causing additionally incomplete pathways for valine and isoleucine. This suggests that both *Methanocorpusculum* lack the capacity for the *de novo* synthesis of many amino acids by typical pathways, and either use uncharacterised schemes for their synthesis or are auxotrophic. As host-associated methanogens, this lack of amino acid biosynthesis suggests the organisms are like provided a constant source of amino acids through the host diet, with both species also containing annotated amino acid transporters.

Functions potentially involved with surface decoration, motility, and adhesion

Both *Methanocorpusculum* sp. MG and CW153 contain a similar number of predicted transmembrane proteins with 536 (24.79%) and 512 (25.09%, Table 4.3). These genes were associated with membrane transport, glycosylation, and membrane structure, although a large proportion are annotated as “hypothetical”. Both genomes encode copies of *flaG/flaF* family flagellin (archaellin), flagellar protein *flaJ*, preflagellin peptidase *flaK*, and other flagellin-like annotated proteins. Both genomes also contain type IV secretory pathway VirB2 component, associated with the production of pili, which facilitates intercellular contacts (Alvarez-Martinez & Christie, 2009). The *Methanocorpusculum* sp. MG genome is predicted to contain additional copies of *flaG/flaF* family

flagellin (archaellin), and flagellin-like protein, as well as pilus assembly pilin *flp* not encoded by *Methanocorpusculum* sp. CW153. Both genomes also contained predicted transmembrane beta-lactamase superfamily and multiple antibiotic resistance proteins, along with several copies of *emrB/qacA* subfamily drug resistance transporters. This suggests both host-associated *Methanocorpusculum* encode for flagellin-like structures, along with other transmembrane and transport associated proteins that likely represent adaptations to their host environment and adherence to host cells.

Collectively, the functions predicted from the genome annotations outlined above suggest both host-associated *Methanocorpusculum* spp. use the canonical scheme of hydrogenotrophic methanogenesis and are likely unable to utilise short-chain alcohols in CO_2-dependent methanogenesis. However, they do contain several genes associated with methanogenesis that currently have limited functional characterisation. Additionally, these isolates require exogenous sources of amino acids, and encode for flagellin-like structures and transport/resistance proteins that likely represent host adaptions. Additionally, although both are able to fix nitrogen, *Methanocorpusculum* sp. MG encodes for more diverse pathways of nitrogen fixation compared to *Methanocorpusculum* sp. CW153, and thus may fix nitrogen at a greater efficiency.

4.3.5 Light microscopy and TEM of novel *Methanocorpusculum* isolates

Methanocorpusculum sp. MG and CW153 stained Gram-negative and presented as pleomorphic cells, ~0.5 to 1.5 µm in diameter. Viable *Methanocorpusculum* sp. MG and CW153 cells are auto-fluorescent at 420 nm, due to the presence of the reduced form of cofactor F_{420} that coordinates the proton and electron transfer during C1 reduction (Figure 4.6, (Cheeseman et al., 1972; Graham & White, 2002). A similar irregular coccoidal morphology has been described for *Methanocorpusculum parvum* and *Methanocorpusculum labreanum*, along with similar cell diameters (Zellner et al., 1987; Zhao et al., 1989). The TEM images of *Methanocorpusculum* sp. MG and CW153 show a clearer picture of the pleomorphology shown in Figure 4.6, with both strains found to possess a relatively thin cell wall and singular membrane that most likely explains the Gram-negative staining of both strains (Figure 4.7). Neither strain produced any obvious capsule-like structure, although *Methanocorpusculum* sp. MG perhaps produces a fine polysaccharide layer, as shown in Figure 4.7B. Interestingly, neither the *Methanocorpusculum* sp. CW153 or MG cells show any evidence for the formation of flagellar or pilin-like structures, despite the genomes of both isolates containing multiple genes annotated as encoding these functions (Figure 4.7). Although some movement can be observed using light microscopy, it is difficult to attribute this solely to the presence of flagellar or pilin-like

structures. It is possible the flagellar or pilin-like structures are only expressed under stress or nutrient deficient conditions that are not induced by the nutrient rich growth medium.

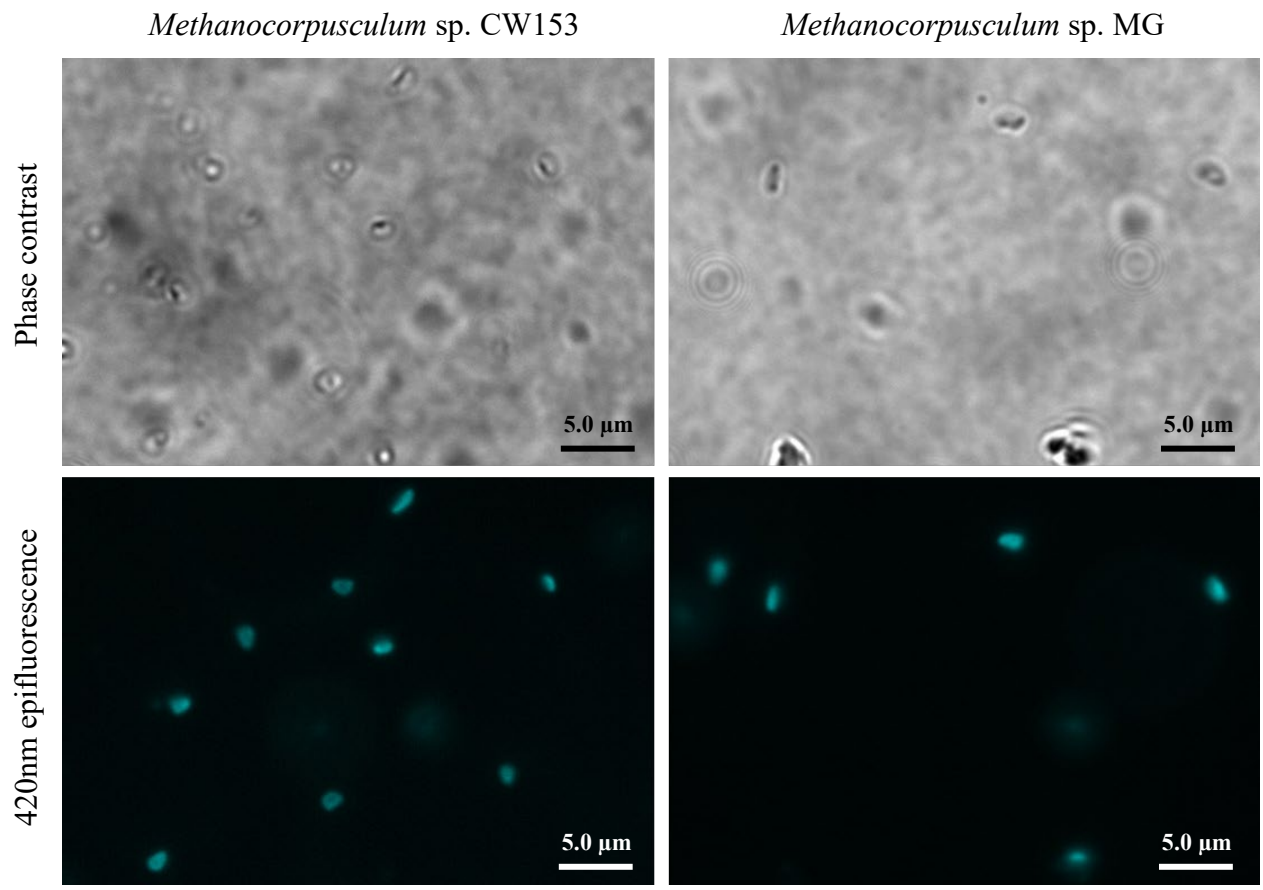


Figure 4.6. Micrographs of *Methanocorpusculum* sp. CW153 and MG. Samples of Methanocorpusculum sp. CW153 and MG cultures were imaged under phase contrast and epifluorescence at 420 nm to observe auto-fluorescence. *Methanocorpusculum* sp. MG and CW153 were both auto-fluorescent.

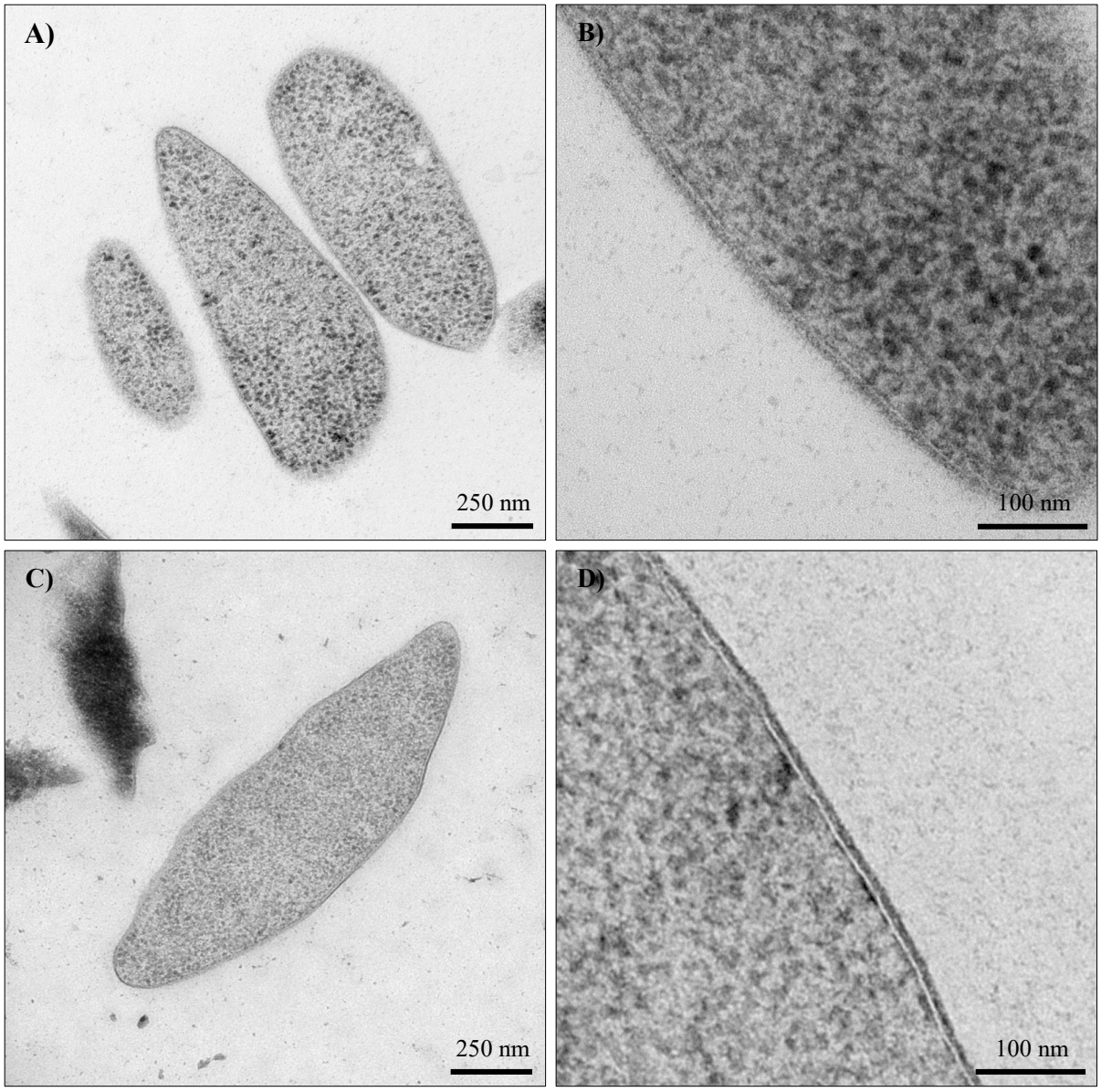


Figure 4.7. Transmission electron micrographs (TEM) of *Methanocorpusculum* sp. CW153 and MG. A-B) *Methanocorpusculum* sp. MG, **C-D)** *Methanocorpusculum* sp. CW153. TEM was conducted by Rick Webb at the University of Queensland Centre for Microscopy and Microanalysis, as per the methods. Both species showed pleomorphic cells with a simple cell wall and no capsular structures.

4.3.6 Bioinformatic validation of host-associated *Methanocorpusculaceae*

The methanogen MAGs I have produced from the publicly available faecal metagenome datasets of various animal species are shown in Figure 4.8. Collectively, I produced 739 methanogen MAGs from the 97 datasets comprised of 5,392 individual metagenomes, which represent no less than 31 host species from 21 countries (Table 6.15). I then augmented this collection of MAGs with published methanogen MAGs and genomes from cultured isolates, which produced a methanogen database containing 2246 genomes from 64 animal species. As expected, genomes affiliated with the family *Methanobacteriaceae* were numerically predominant, representing 72% of the total. The family *Methanomethylophilaceae* ranked second at ~18% and surprisingly, the family *Methanocorpusculaceae* was the third most dominant family at ~6.5%. *Methanocorpusculum* MAGs

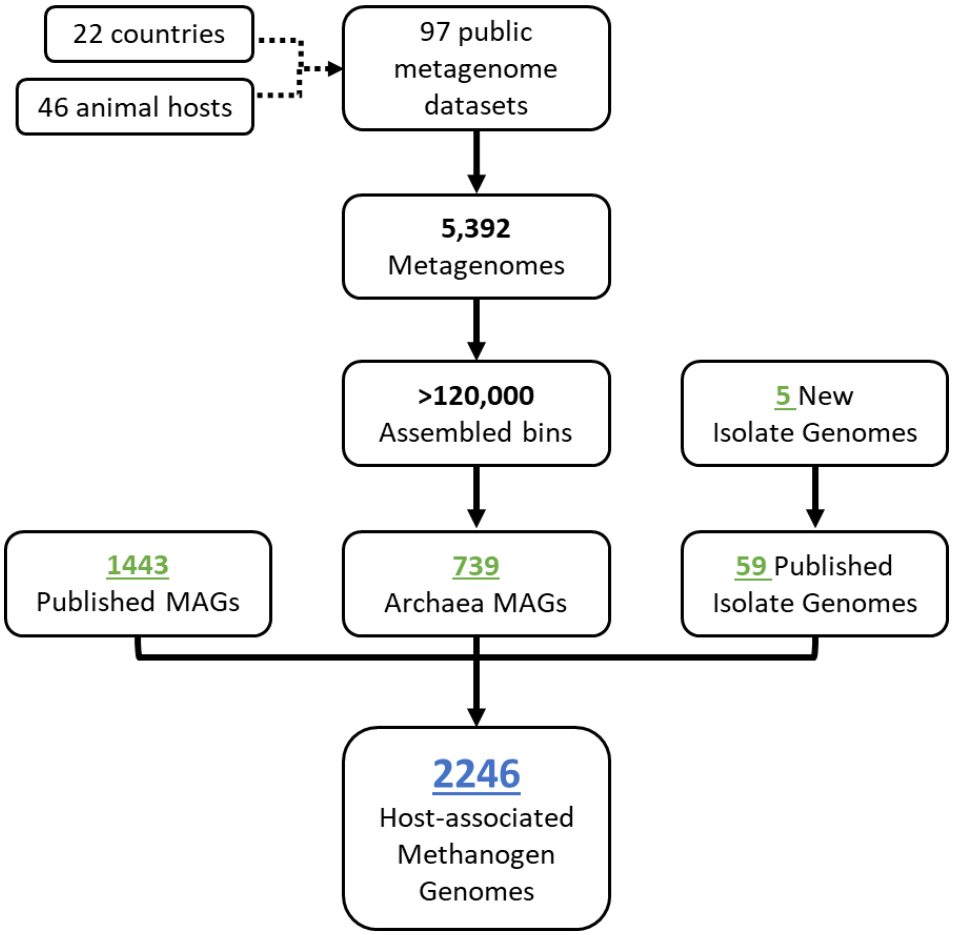


Figure 4.8. Flow chart showing the generation of a host-associated methanogen database from diverse animal species. 5,392 metagenomes from 97 metagenome datasets were used to recover methanogen MAGs, as per the methods. 739 methanogen MAGs (≥50% completeness, ≤10% contamination) were successfully recovered from 35 host species. Combined with 1443 published methanogen MAGs and 59 isolate genomes, a host-associated methanogen database was constructed with 2246 total genomes.

were successfully recovered from 20/97 datasets (Table 4.3), suggesting that the newly identified family of gut methanogens is prevalent in the host-associated microbiome. Hence, the two cultured representatives are a significant advance to support the biological characterisation of this previously unrecognised group of host-associated methanogens. To that end, and for my thesis, I have decided to focus on the *Methanocorpusculum* MAGs and isolate genomes, rather than the entire database.

A total of 130 MAGs assigned to *Methanocorpusculaceae* were successfully recovered from publicly available MGS datasets, with 24 MAGs being high-quality (HQ). These recovered MAGs were then combined with 10 MAGs produced from Southern hairy-nose wombats and mahogany gliders by Dr Rochelle Soo, four human-associated MAGs recovered by Nayfach et al. (2019), five

Table 4.3. Metagenomic datasets used to recovery predicted *Methanocorpusculaceae* MAGs from animal hosts.

Study	BioProject	PMID	Animal/s	MAG Prefix ID	Samples Analysed	MAGs Recovered
-	PRJNA590977	-	Horse	PRJNA590977	5	5
-	PRJNA545600	-	Elephant	PRJNA545600	1	2
-	PRJNA545601	-	Elephant	PRJNA545601	1	1
-	PRJNA545606	-	Elephant	PRJNA545606	1	2
Cao et al. (2020)	PRJNA556790, PRJNA563508	32122398	Bird, human	CaoJ	158	3
-	PRJNA427653	-	Cow, Sheep, Goat, Horse, Human, Soil	PRJNA427653	81	13
Doster et al. (2018)	PRJNA309291	30105011	Bovine	DosterE	61	3
Gibson et al. (2019)	PRJNA532626	31138833	Rhinoceros	GibsonKM	25	9
Hou et al. (2016)	PRJNA340908	27876778	Chicken	HouQ	29	9
Ilmberger et al. (2014)	PRJNA240141	25208077	Elephant	IlmbergerN	2	1
Li et al. (2019)	PRJNA411766	31309006	Sperm Whale	LiC	1	1
Lim et al. (2020)	PRJEB32496	-	Bovine, swine	LimSK	77	6
-	PRJEB23356	-	Chicken	PRJEB23356	651	55
-	PRJNA293646	-	Chicken, Pig, bovine	PRJNA293646	13	1
Salgado-Flores et al. (2019)	PRJNA450906	30856229	Ptarmigan	SalgadoFloresA	17	6
H. Wang et al. (2019)	PRJNA483083	30800107	Rhesus monkey	WangH	16	2
Zaheer et al. (2019)	PRJNA529711, PRJNA420682	29651035	Bovine, environmental	ZaheerR	41	7
Rovira Sanz (2017)	PRJNA379303	-	Bovine	SanzPR	96	4

environmentally associated MAGs recovered by Parks et al. (2017), one MAG produced from a wombat by Shiffman et al. (2017), and eight other MAGs identified as *Methanocorpusculaceae* on the NCBI genome database (Table 6.16). These 161 *Methanocorpusculum* genomes were isolated or recovered from 13 geographical locations (Figure 4.9). The host-associated MAGs showed a wider geographical range, though the majority were recovered from samples originating from India and China. The Chinese samples were represented by several studies, but all MAGs recovered from Indian samples were derived from a single study on the metagenome of chickens (PRJEB23356), which produced 55 *Methanocorpusculum* MAGs (Figure 4.10). The Australian genomes, including the two novel isolates *Methanocorpusculum* sp. MG and CW153, represent the third largest geographical group.

The phylogenetic analysis shown in Figure 4.11 suggests there are 12 MAGs identified no further than family_*Methanocorpusculaceae* and were recovered from horses, rhinoceros, and elephants. However, none of these MAGs met the threshold for HQ genomes, which makes these findings inconclusive. The phylogenetic analysis of *Methanocorpusculaceae* archaeon UBA456 and *Methanocorpusculaceae* archaeon WOFA02 shows both genomes are outliers to the *Methanocorpusculum* and are more likely representatives of *Methanocalculus***.** Almost 80% of the MAGs could be identified as genus_*Methanocorpusculum* and as Figure 4.11 shows, the host-associated *Methanocorpusculum* spp. show substantial diversity. There appears to be no less than

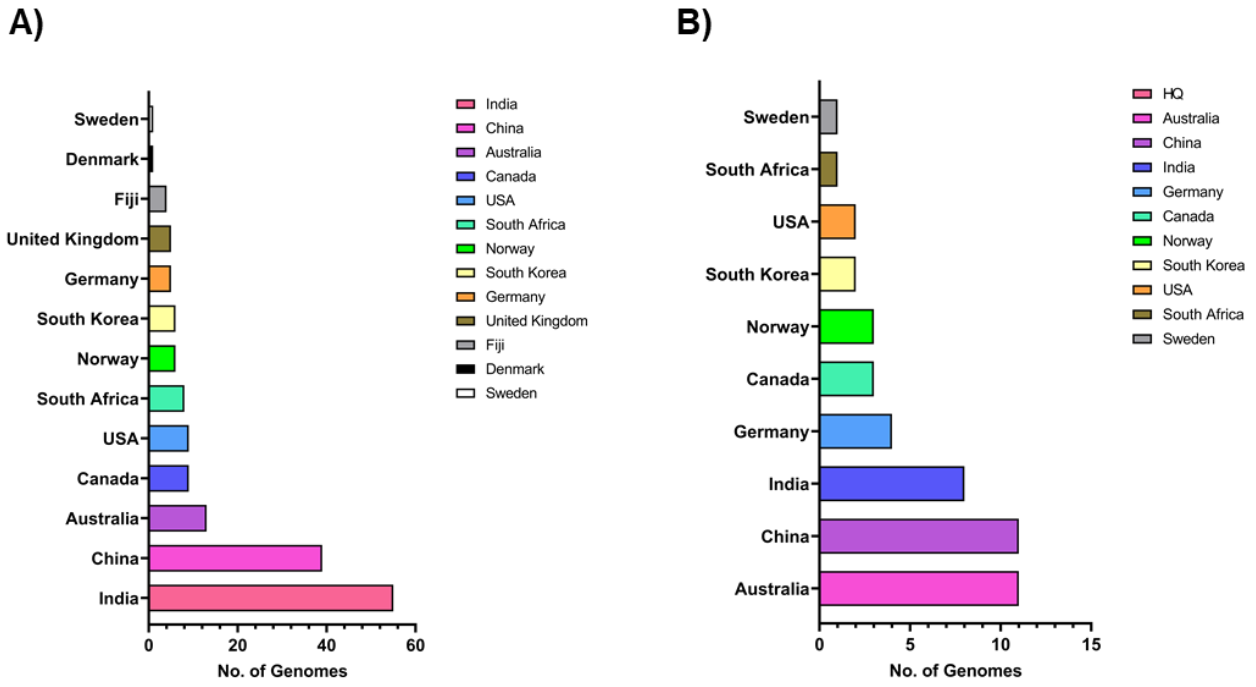


Figure 4.9. Geographical distribution of *Methanocorpusculaceae* MAGs and isolate genomes. **A)** All (≥50% completeness, ≤10% contamination) MAGs, **B)** HQ (≥90% completeness, ≤5% contamination) MAGs.

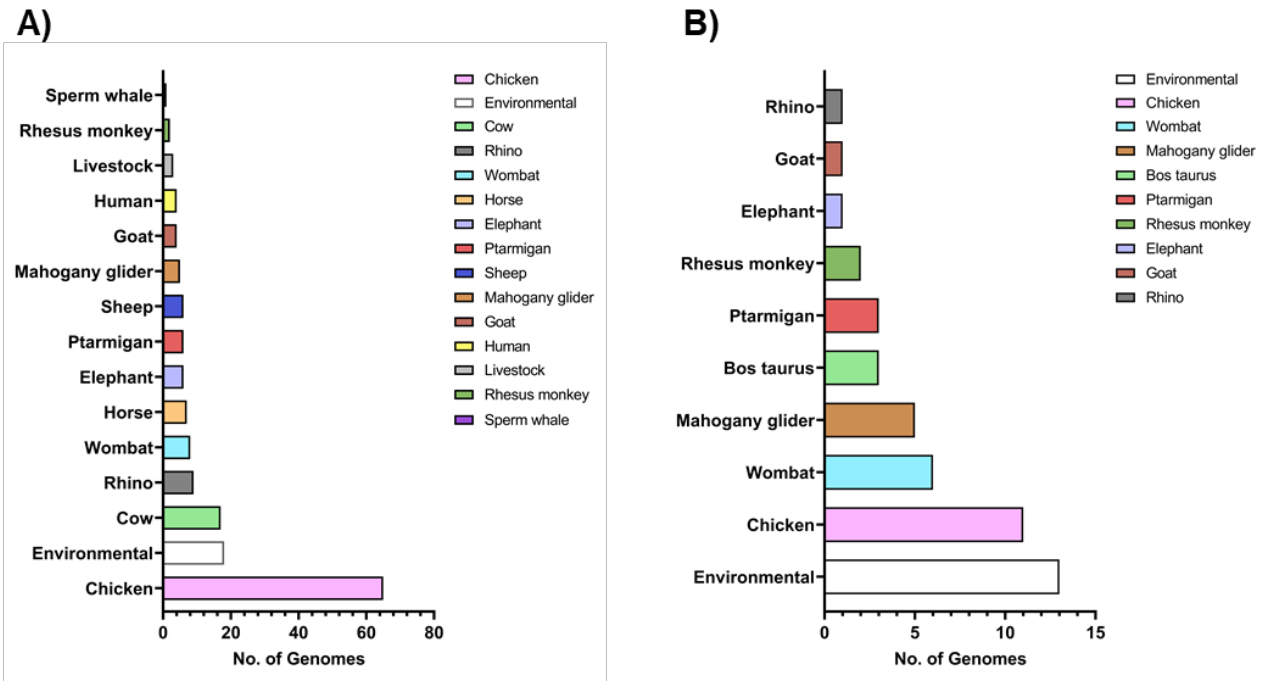


Figure 4.10. Host and environmental distribution of *Methanocorpusculaceae* MAGs and isolate genomes. **A)** All (≥50% completeness, ≤10% contamination) MAGs, **B)** HQ (≥90% completeness, ≤5% contamination) MAGs.

four environment-associated (Env) clades, and 18 host-associated (HA) clades, of which 17 HA clades are novel. The MAGs recovered from rhinoceros, elephant, and horse samples contain the greatest diversity, with MAGs from these animals assigned to nine clades (Figure 4.11; HA clade 1-9). Interestingly, two MAGs from rhinoceros produce an outlying clade compared to all other *Methanocorpusculum* genomes (Figure 4.11; HA clade 1). One MQ MAG (LiC_2019_SRR6192929_bin.122) recovered from a sperm whale was phylogenetically distinct but grouped closest with HA clade 3, containing MAGs derived from elephant and rhinoceros. The MAGs recovered from ruminant animals (i.e. sheep, cows, and goats) were assigned to only three clades, along with MAGs recovered from samples of undefined 'livestock' (Cao et al., 2020). Additionally, the *Methanocorpusculum* MAGs produced from humans by Nayfach et al. (2019) also clustered with the ruminant MAGs, in HA clade 12. The *Methanocorpusculum* MAGs recovered from ptarmigan produced its own distinct clade (Figure 4.11; HA clade 14), which could be attributed to these *Methanocorpusculum* being recovered from a non-*Mammalian* host. The MAGs recovered from Australian marsupials separated into two distinct clades: those produced from mahogany gliders assigned to HA clade 15 and those produced from wombats in HA clade 17, except for one assigned to HA clade 15. Notably, the genomes from strains CW153 and MG were assigned to HA clades 15 and 17, respectively. Interestingly, the two MAGs produced from the rhesus macaque datasets (WangH_SRR7619055_bin.13, WangH_SRR7619059_bin.15) were also assigned to HA clade 17

and two MAGs recovered from elephants produced a distinct clade that branches off HA clade 17 (HA clade 16). HA Clade 18 represents the most dominant clade, with 65 MAGs recovered from chickens, likely due to the over representation of chicken metagenomes (53%) in the non-human metagenomes used to recover MAGs. Interestingly, the chicken MAGs cluster closer with the marsupial clades than the ptarmigan in HA clade 14, despite both clades representing MAGs recovered from avian hosts.

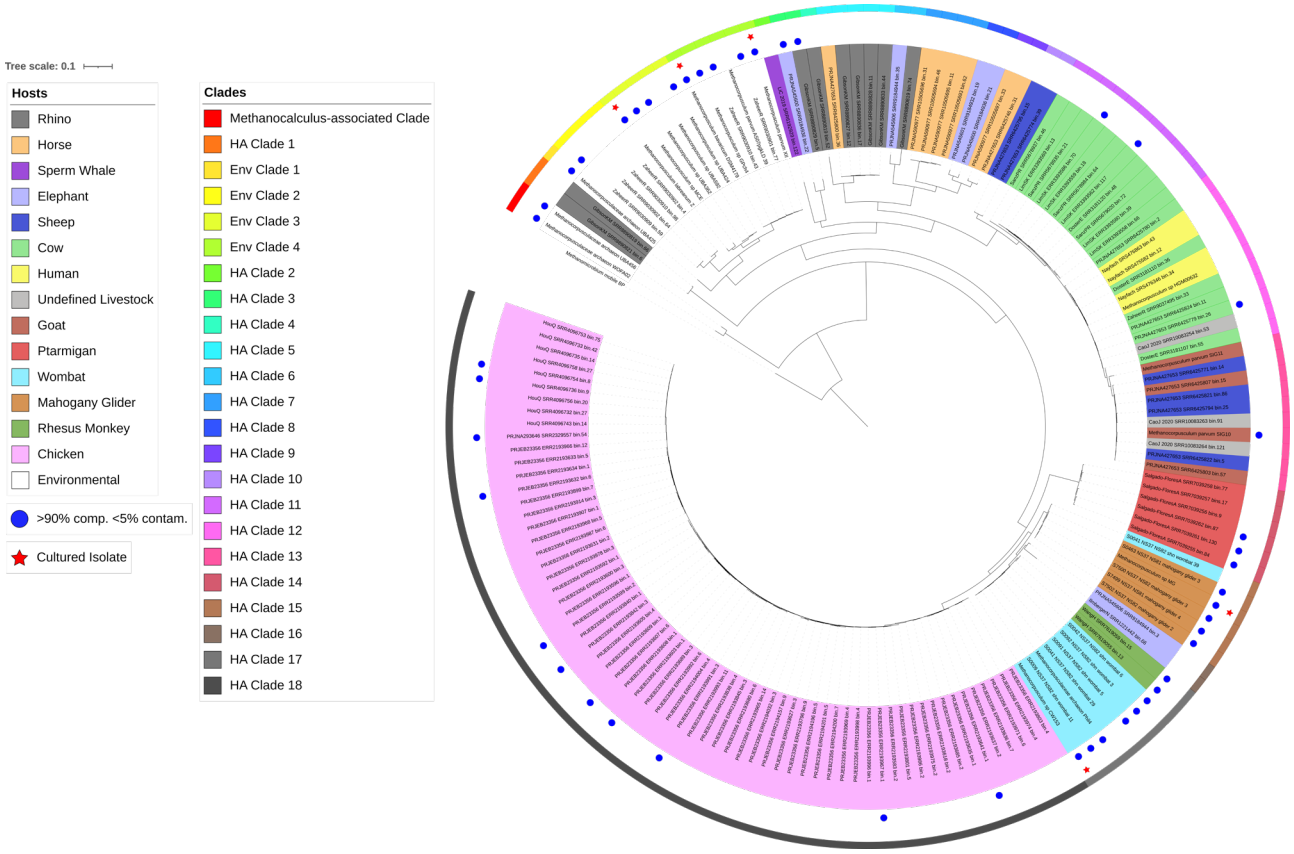


Figure 4.11. Phylogenetic distribution of *Methanocorpusculaceae* MAGs and isolate genomes. Concatenated archaeal marker gene files were produced using GTDB-tk (v.1.3.0), with *Methanomicrobium mobile* BP used as the outgroup. Phylogeny was inferred using FastTree (v2.1.10) and visualisation by iToL (https://itol.embl.de/). MAGs and isolate genomes of ≥50% completeness and ≤10% contamination were included, and HQ MAGs identified by blue circles. Cultured isolates were identified by a red star. All MAGs and isolate genomes from environmental sources were identified as 'Environmental' under host description. MAGs and isolate genomes clustered into 18 host-associated (HA) clades, four environmental (Env) clades and one *Methanocalculus*-associated clade.

4.3.7 Clade-specific variations in *Methanocorpusculum* genome content and metabolic potential

For these analyses, only the 44 HQ *Methanocorpusculum*-affiliated MAGs and isolate genomes were used, and represent the four Env clades, and 6/18 HA clades. The genomes from the cultured isolates represent 2/4 Env clades, and 2/6 HA clades included in these analyses (Figure 4.12). These criteria excluded all the MAGs recovered from horses, sperm whales, sheep, undefined livestock, and humans. The collective comparative matrix of these genomes by average nucleotide identity is shown in Figure 6.4 but the individual genomes were sufficiently dissimilar to all others (<70%) for any accurate

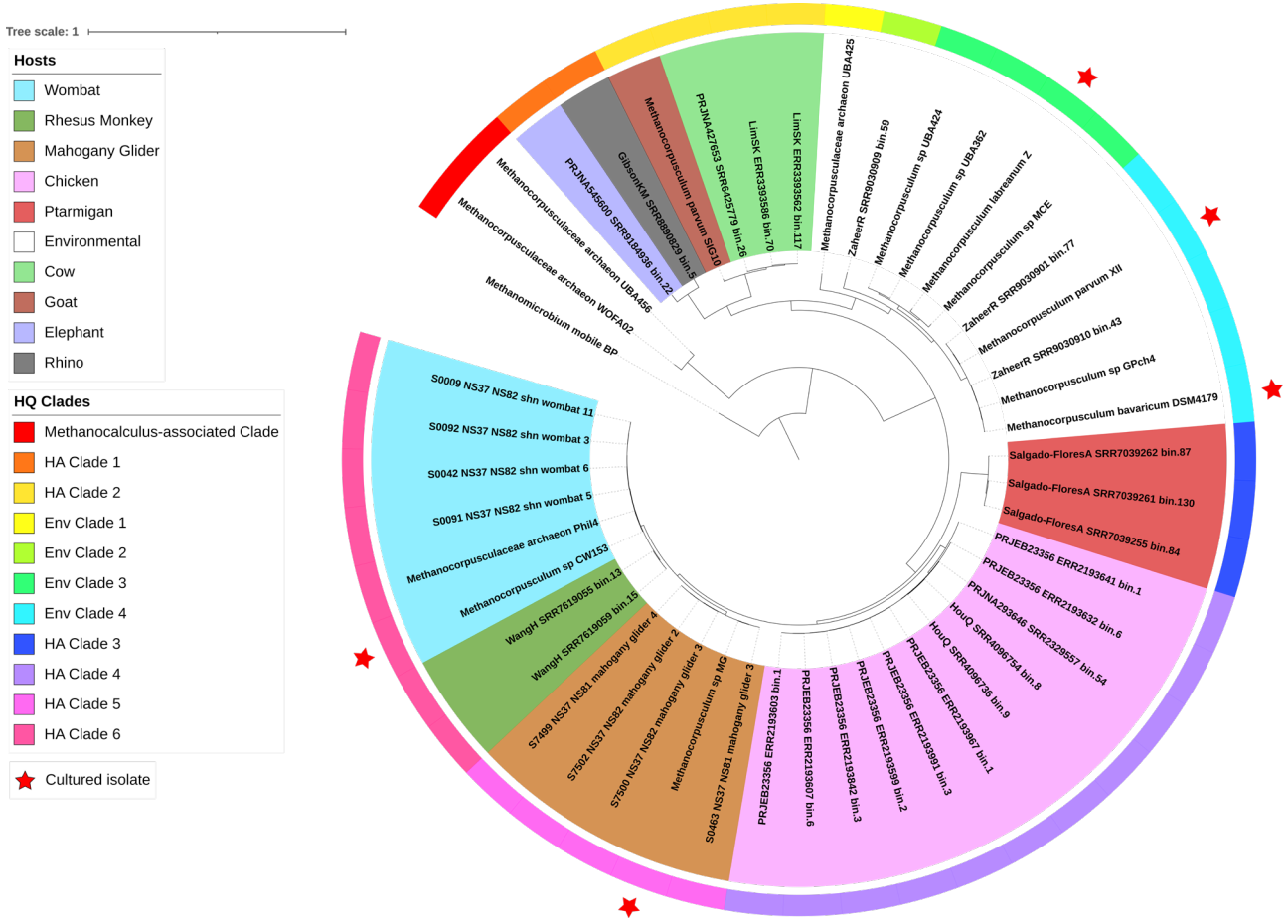


Figure 4.12. Phylogenetic distribution of high-quality (HQ) *Methanocorpusculaceae* MAGs and isolate genomes. Concatenated archaeal marker gene files were produced using GTDB-tk (v.1.3.0), with *Methanomicrobium mobile* BP used to root the tree. Phylogeny was inferred using FastTree (v2.1.10) visualisation by iToL (https://itol.embl.de/). MAGs and isolate genomes of ≥50% completeness and ≤10% contamination were included, and HQ genomes were identified by blue circles. Cultured isolates were identified by a red star. All MAGs and isolate genomes from environmental sources were identified as 'Environmental'. MAGs and isolate genomes clustered into six host-associated (HA) clades, four environmental (Env) clades and one *Methanocalculus*-associated clade.

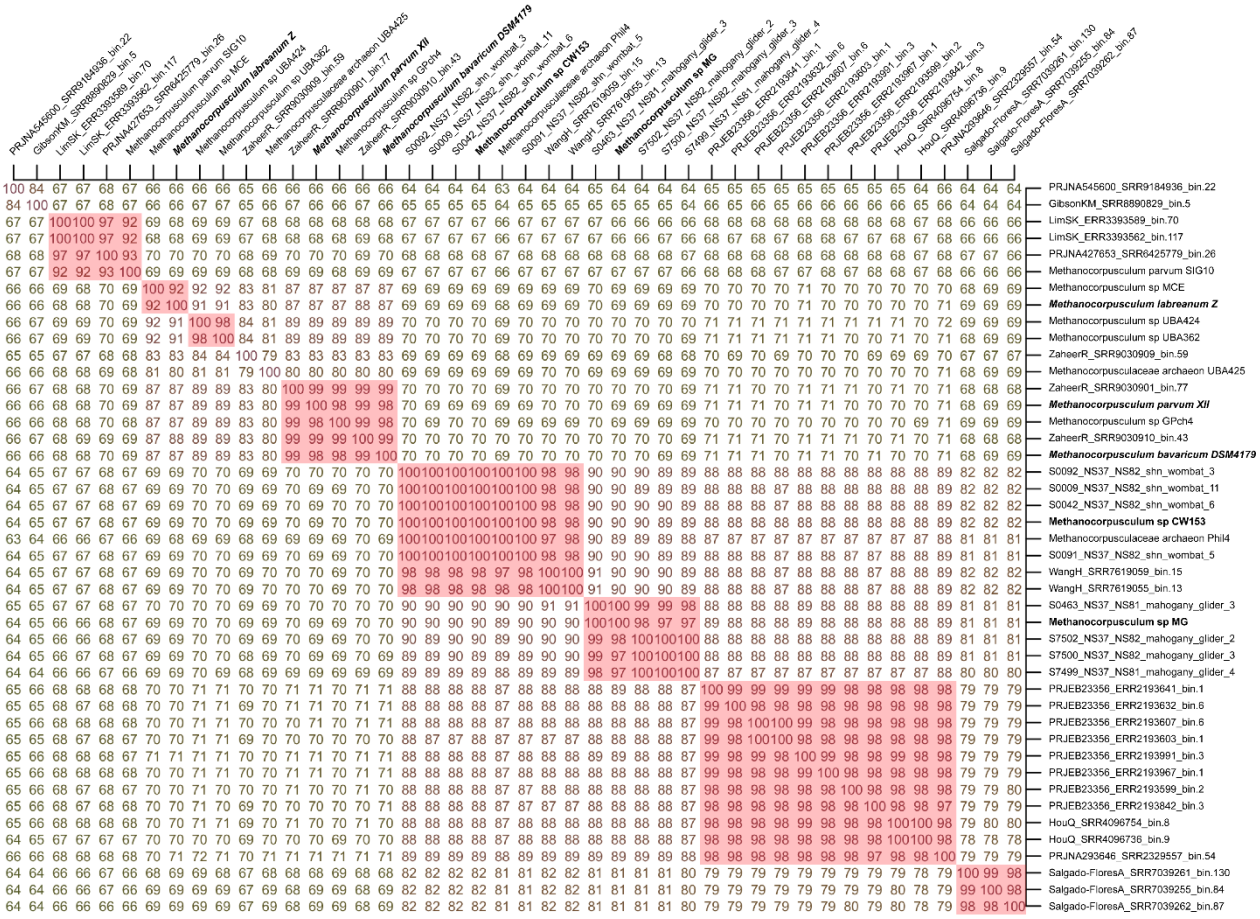


Figure 4.13. Average amino acid identity (AAI) of *Methanocorpusculum* MAGs and isolate genomes. The values and heatmap were generated using the Kostas lab online AAI calculator (http://enve-omics.ce.gatech.edu/aai/). The genomes of cultured isolates have been shown in bold. AAI shows the presence of at least eight *Methanocorpusculum* phylogroups, with four genomes clustering alone and potentially represent additional phylogroups.

categorisation. Therefore, the MAGs and genomes were compared to each other by average amino acid identity (AAI). These analyses suggest that there are no less than eight distinct clades, with five HA clades and three Env clades (Figure 4.13). The elephant and rhino MAGs (PRJNA545600_SRR9184936_bin.22 and GibsonKM_SRR8890829_bin.5, respectively) clustered closest with each other (84%) and only showed 68% maximum identity to other *Methanocorpusculum*, confirming that they represent outliers to the other MAGs and likely represent a novel genus. Similarly, the Env MAGs ZaheerR_SRR9030909_bin.59 and Methanocorpusculaceae_archaeon_UBA425 did not cluster with any other Env clade. The differentiation of species by AAI is typically assessed by a ~95% cut-off, meaning those with less than ~95% similarity represent different species, with sequentially smaller percentages representing genus and family classifications (Konstantinidis & Tiedje, 2005). As such, the lower values of 64-70% found here warrant the consideration that the MAGs and isolates genomes of environmental and

host-associated origins, as well as elephant and rhino MAGs represent, form deep lineages within the family *Methanocorpusculaceae*.

Analysis of the *Methanocorpusculum* genomes using the Bacterial Pan-Genome Analysis (BPGA) software showed the elephant and rhino MAGs to contain the largest number of unique genes at 18% and 17% (289 and 260 genes, respectively) again suggesting that they represent a divergent clade of *Methanocorpusculum* (Table 6.17). Notably, there is still a progressive increase in the number of unique genes added to the pangenome from each MAG/genome introduced into this analysis and resulted in 7263 genes contributing to the *Methanocorpusculum* "pangenome" (Figure 4.14). Only 180 genes (~2.5%) are considered as core genome content, which is shared by all. However, when the Env and HA MAGs/genomes were considered separately, the core genome increased to 890 genes

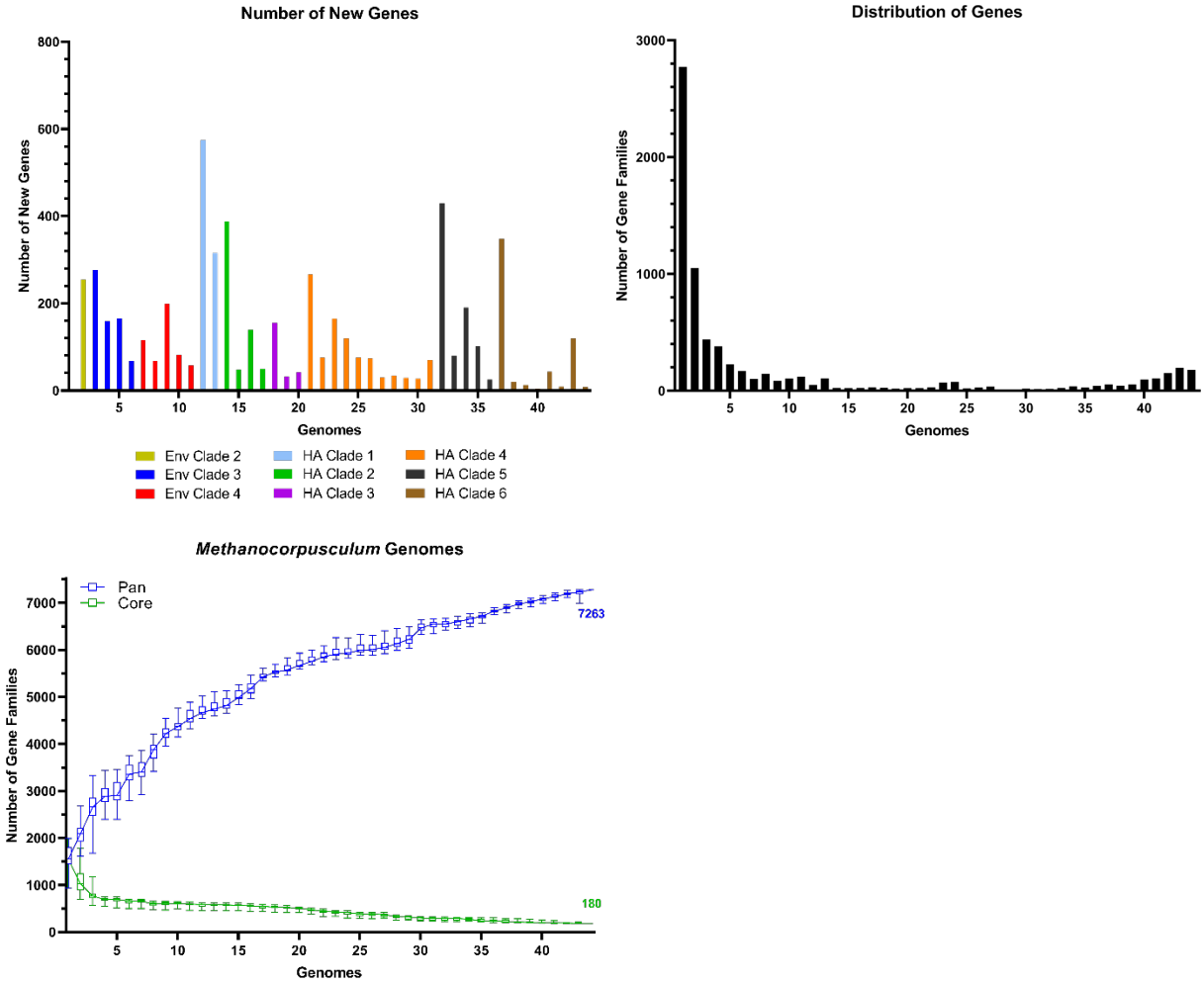


Figure 4.14. Comparative analysis of the core and pan genome of *Methanocorpusculum*. Core and pan genome analysis was conducted using the Bacterial Pan Genome Analysis software (Chaudhari et al., 2016), as per the methods. An increase in the presence of new genes was seen for each predicted HA or Env clade. 7293 genes were found within the pan genome, with only 180 core genes across all *Methanocorpusculum* genomes.

(~29% of the pangenome) for the Env genomes, whereas only 224 genes (~4% of the pangenome) were considered core to the HA genomes (Figure 4.15). To account for potential sampling bias, the HA genomes were also analysed at 10 genomes, which contained 533 core gene and maintained the significantly smaller core genome size ($P<0.0001$). Unsurprisingly, the core genome predicted for each clade are larger, ranging in size from 725 genes (HA clade 4) to 1441 genes (HA clade 6). However, these numbers are likely to change because only the pangenome from those MAGs/genomes assigned to HA clades 3 is described as "almost closed", the remaining clade-specific Env and HA pangenomes are defined as "still open but may be closed soon". Despite their incompleteness using the available data, there is a high degree of functional diversity inherent to each *Methanocorpusculum* clade of host- and environmental origin. As such, the results also provide strong evidence that these lineages are habitat-specific (i.e. host-associated and environmental origins) rather than something encountered by chance.

Next, I examined the COG and KEGG-Orthology (KO) based profiles for each *Methanocorpusculum* clade separated into their respective core, accessory, and unique genes. As expected, the distribution of genes across the different COG categories is relatively consistent for the different *Methanocorpusculum* clades (Figure 4.17). For all clades, the COG categories for general functional prediction only (R) and unknown functions (S) were the most abundant (Figure 4.17) for the core and accessory genes. In relative terms, the HA clade 3 pangenome has a smaller number of genes assigned to R and S categories, which is likely attributed to their smaller average genome size and the fewer relative number of predicted genes for genomes within this clade. The housekeeping genes assigned to the core genome, such as those encoding functions involved with translation, ribosomal structure,

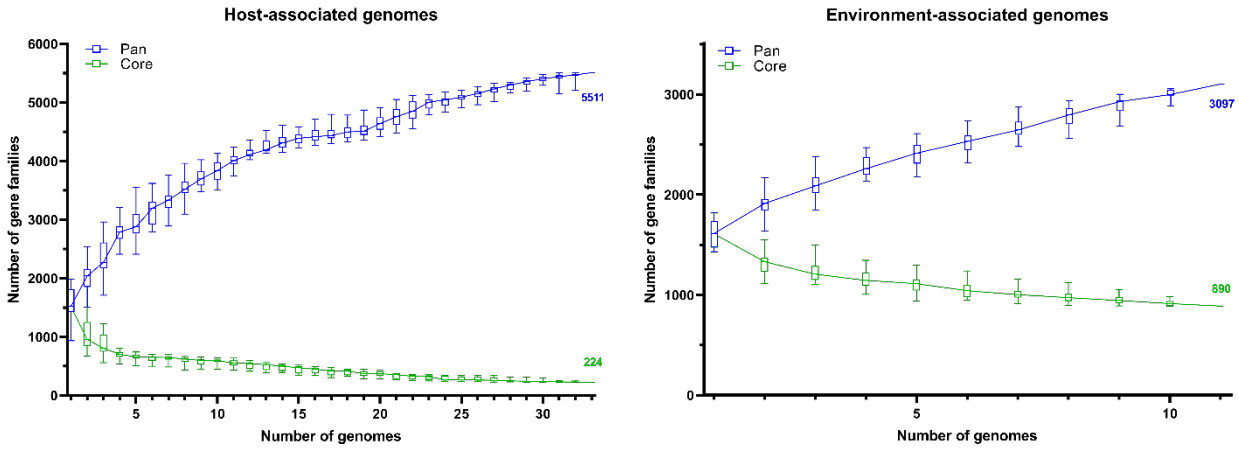


Figure 4.15. Core and pan genome plots for host-associated (HA) and environmental (Env) *Methanocorpusculum* genomes. Core and pan genome analysis was conducted using the Bacterial Pan Genome Analysis software (Chaudhari et al., 2016). Env genomes contained a higher number of core genes at 890 compared to HA genomes which only shared 224 core genes.

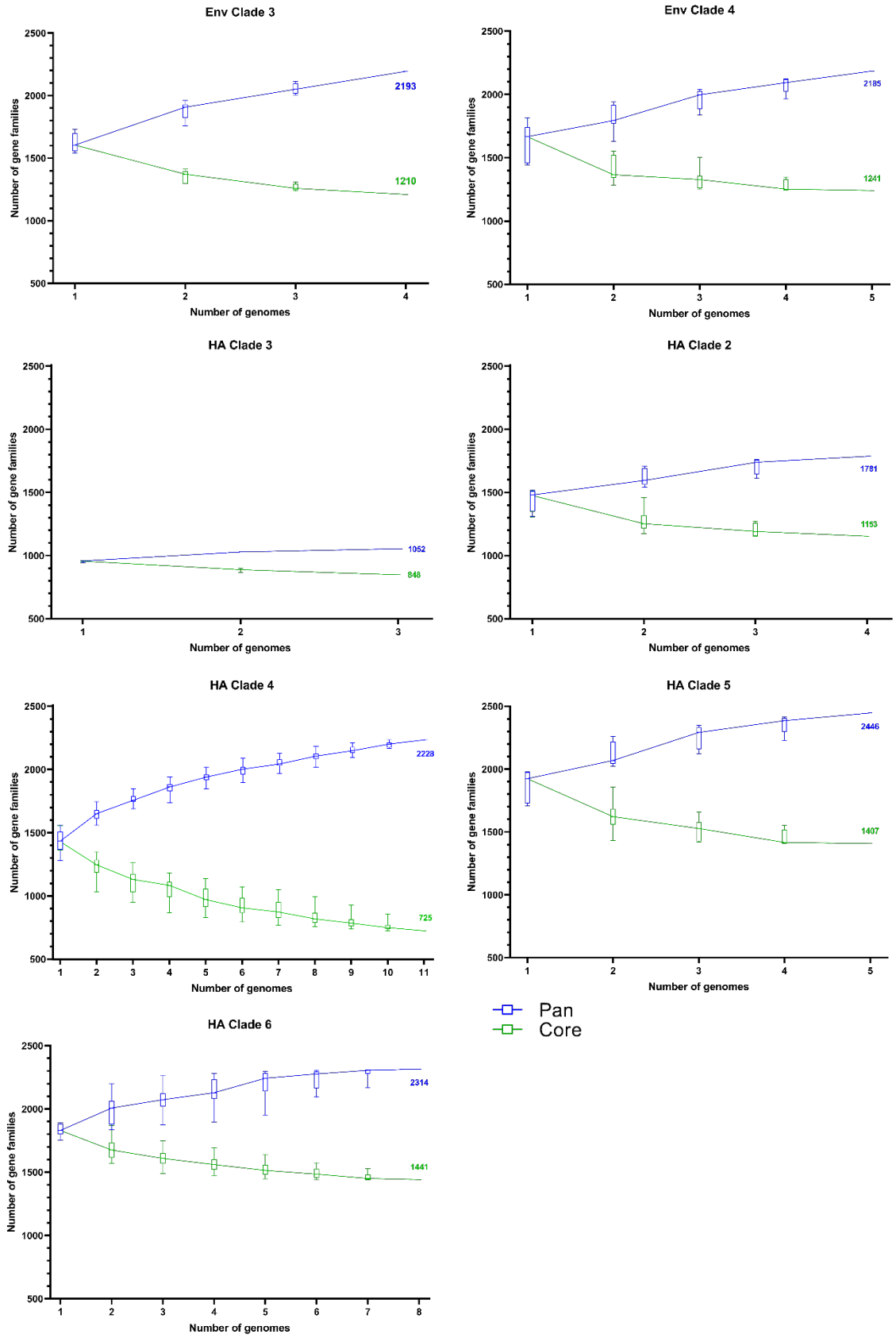


Figure 4.16. Core and pan genome plots for host-associated (HA) and environmental (Env) *Methanocorpusculum* clades. Core and pan genome analysis was conducted using the Bacterial Pan Genome Analysis software (Chaudhari et al., 2016). The marsupial-associated HA clades 5 and 6 showed the highest number of core genes and HA clade 4 showed the lowest.

and biogenesis (J); energy production and conversion (C); and amino acid transport and metabolism (E) had similar proportional representations across the clades, except for the pangenome for HA clade 3, which shows a higher relative percentage of COGs annotated as replication, recombination and repair (L), and energy production and conversion (C). The COG category profiles for accessory genes were also similar across the clades, with the exception again that HA clade 3 that shows a higher percentage of accessory genes predicted to encode for translation, ribosomal structure, and biogenesis functions (J); and Env clade 3 has a higher relative percentage of accessory genes for amino acid transport and metabolism (E). The most common COG categories for the clade-specific unique genes were predicted to be involved with cell wall/membrane/envelope biogenesis (M); defence mechanisms (V); and replication, recombination, and repair (L). More than 30% of HA clade 3 unique gene COG annotations were assigned to energy production and conversion (C), greater than any other

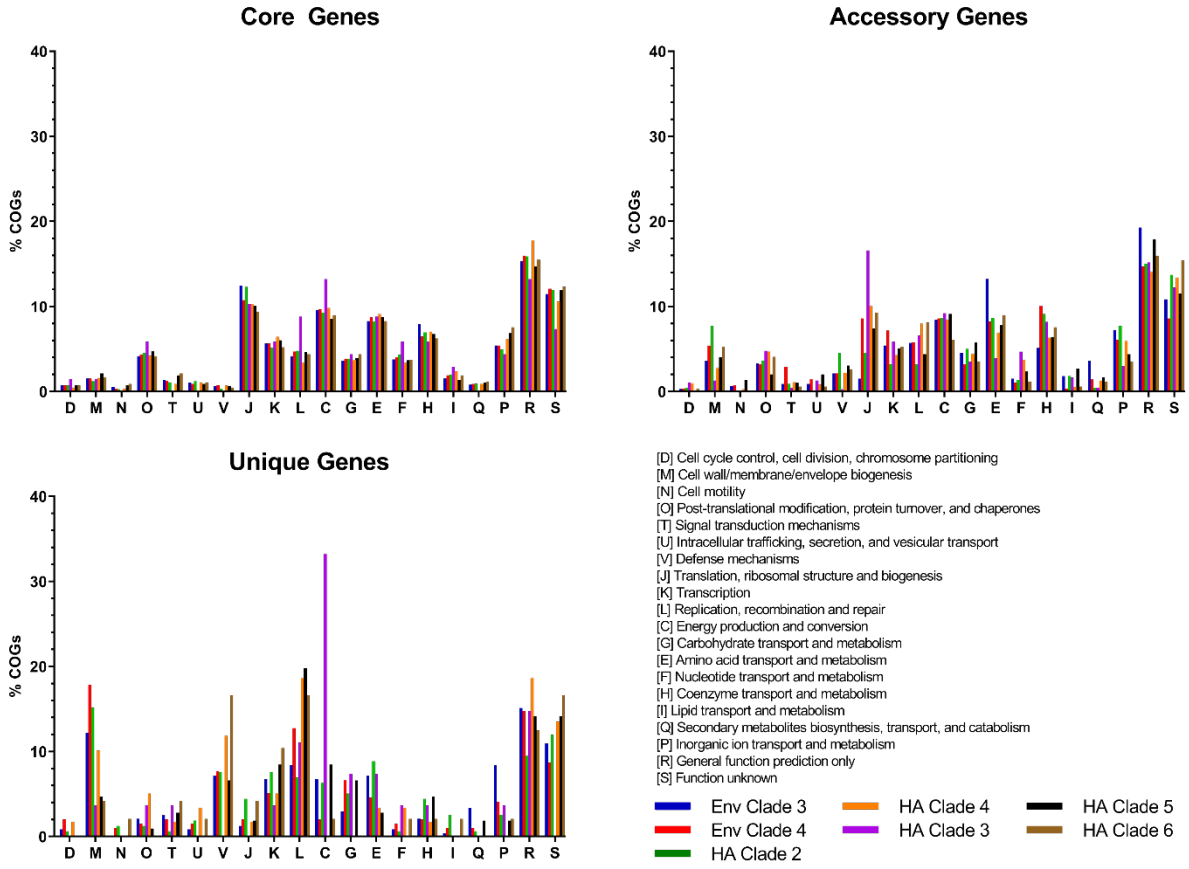


Figure 4.17. COG classification of core, accessory, and core genes for host-associated (HA) and environmental (Env) *Methanocorpusculum* clades. The annotation of COG categories was performed using the Bacterial Pan Genome Analysis software (Chaudhari et al., 2016). Only clades with ≥3 genomes could be compared, thus Env clade 1, 2 and HA clade 1 were excluded. Core and accessory gene COG categories were relatively consistent across the clades, however specific COG categories were differentially observed in the unique genes of the different clades.

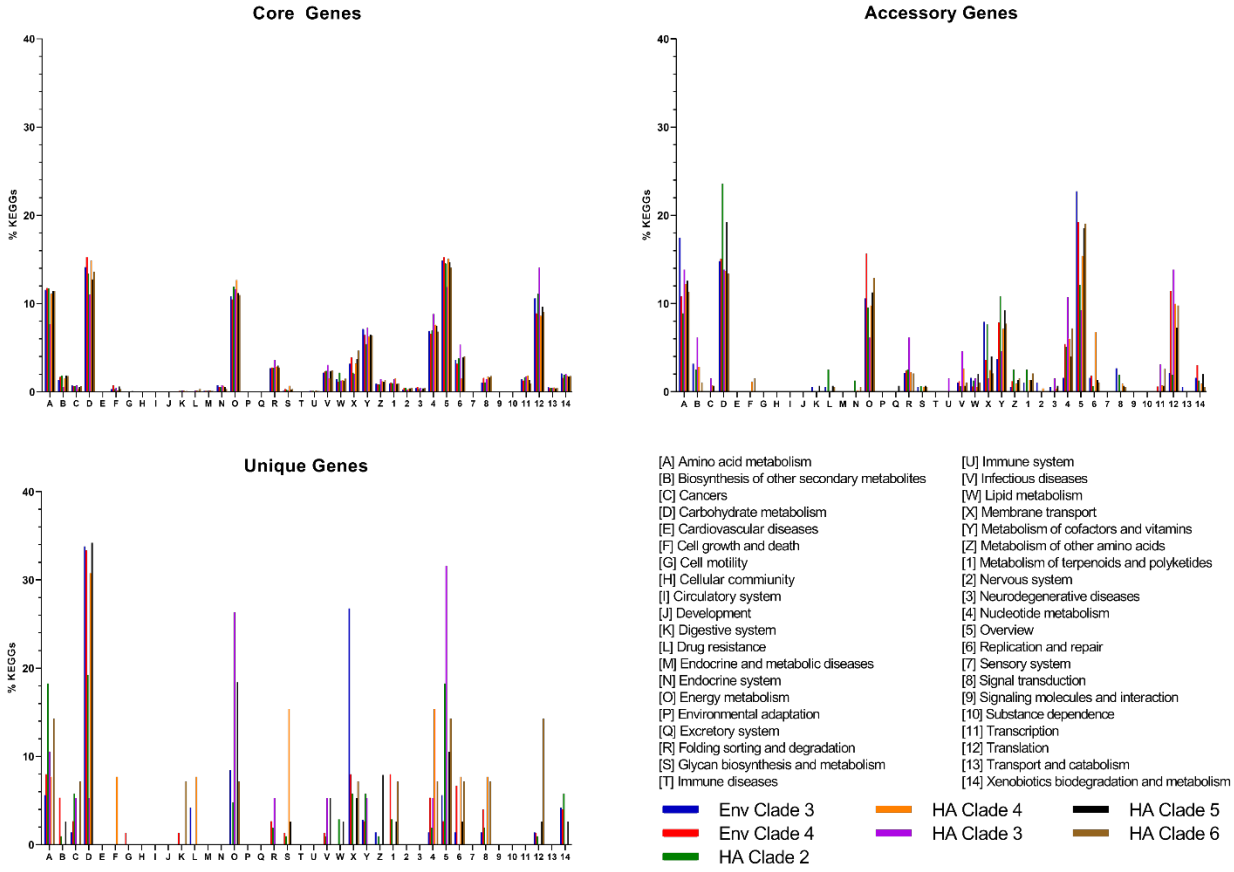


Figure 4.18. KEGG classification of core, accessory, and core genes for host-associated (HA) and environmental (Env) *Methanocorpusculum* clade. KEGG annotation was performed using the Bacterial Pan Genome Analysis software (Chaudhari et al., 2016). Only clades with ≥3 genomes could be compared, thus Env clade 1, 2 and HA clade 1 were excluded. Core and accessory gene KEGG categories were relatively consistent across the clades, however specific KEGG categories were differentially observed in the unique genes of the different clades.

clade. Interestingly, the unique genes of the Env clades, HA clade 2, and HA clade 4 showed a higher percentage of COGs for cell wall/membrane/envelope biogenesis (M). Although HA clades 5 and 6 are both marsupial-associated clades, HA clade 6 showed a higher percentage of COGs for defence mechanisms (V). HA clade 4-6 also showed a comparatively greater percentage for replication, recombination, and repair (L), though it was a dominant category for all clades.

When the KO annotations were used for the clade-specific pangenome analyses (Figure 4.18), amino acid (A), carbohydrate (D), and energy (O) metabolism, as well as general overview (5), and translation (12) were the largest categories represented within the core and accessory genomes of all clades. Interestingly, genes with KO annotations assigned to carbohydrate metabolism (D) represented more than 30% of the unique genes for the pangenomes of Env clade 3 and 4, as well as HA clades 5 and 6. Like the COG-based profiles, the pangenome for HA clade 3 shows a unique

profile, with high percentages of genes assigned to KO categories for energy metabolism (O) and general overview (5). Likewise, the KO profiles for the Env clade 3 pangenome encodes for a greater relative percentage of membrane transport functions (X). The HA clade 4 pangenome was differentiated from the others by the greater percentage of unique genes with KOs predicted to encode functions involved with cell growth and death (F), glycan biosynthesis and metabolism (S), and nucleotide metabolism (4), while unique genes annotated as being involved with translation (12) were greater in HA clade 6.

Collectively, the COG and KO profiles for the clade-specific pangenomes show consistent annotations for the core and accessory genes. However, the unique genes of each clade show substantial variation in COG and KO annotations associated with carbohydrate metabolism, membrane transport and glycan metabolism.

4.3.8 The genome content between M*ethanocorpusculum* MAGs/genomes of environmental and host origin are different.

Based on these results, I then used EnrichM (Boyd et al., 2019) to look at the gene orthologs in the *Methanocorpusculum* clades. As shown in Figure 4.19, there is significant variation between the

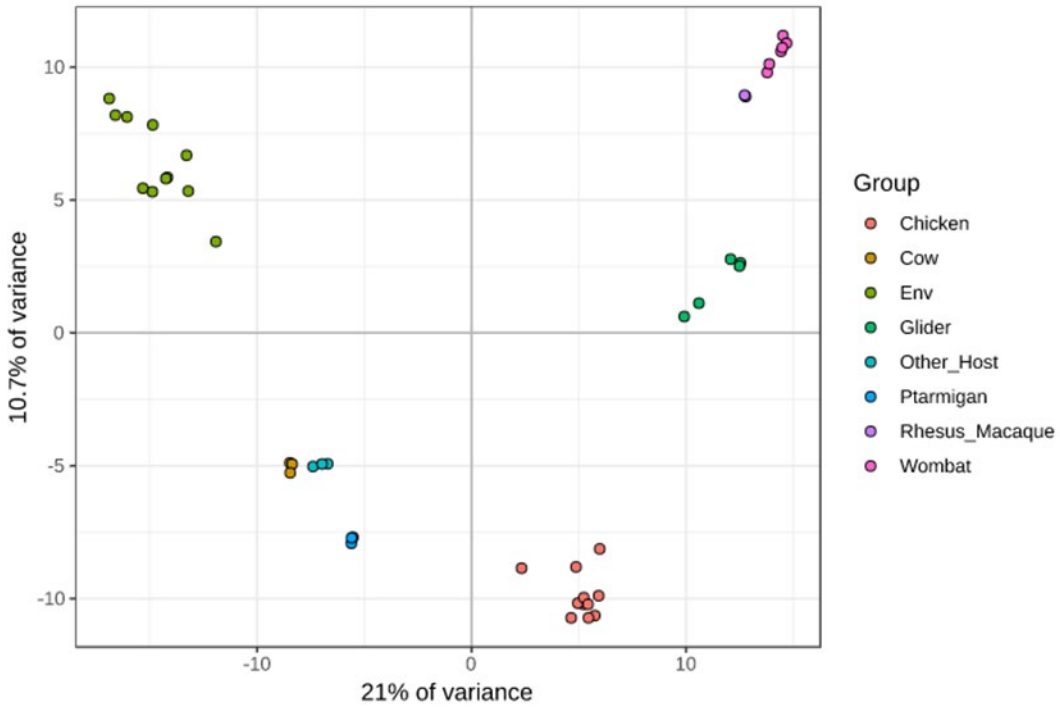


Figure 4.19. Principle component analysis (PCA) plot showing the genetic variance in HQ *Methanocorpusculum* genomes recovered from different environments. The PCA plot was generated by EnrichM (v0.4.15) using the '--orthologs' analysis. Genomes groups were coloured according to the legend. At least two genomes were required for genome groups, as such, the goat, elephant and rhino genomes were combined into the 'Other_Host' group.

different *Methanocorpusculum* clades. Despite the phylogenetic similarity, there was substantial variance in gene orthologs between the marsupial and chicken clades. Interestingly, the ptarmigan MAGs also clustered closely with the MAGs derived from ruminants. There was also a clear differentiation between the wombat and mahogany glider-associated clades.

Based on these results, I then used EnrichM to evaluate whether the KO annotation profiles differ between the MAGs and genomes of different origins (i.e. Env versus HA). The gene counts assigned 105 KO as significantly different between these two groupings by Fisher's Exact test (Figure 4.20A-B), with 35 KO annotations enriched in the HA genomes and 70 enriched in the Env genomes. In contrast, when the KO categories were compared based on gene count per genome by the Mann-Whitney U test (Figure 4.21), 74 KO were significantly different between the Env and HA groups. This is likely due to the Mann-Whitney U test comparing the distribution of genes between the two genome groups, whereas the Fisher's Exact test is a non-parametric test for the independence of *Methanocorpusculum* groups based on the KOs. The Fisher's Exact test showed that KOs associated with amino acid metabolism were most prominent difference between the Env and HA genomes. Specifically, the Env genomes were significantly enriched for genes encoding tryptophan synthase (*trpA/B*), indole-3-glycerol phosphate synthase (*trpC*), anthranilate phosphoribosyltransferase (*trpD*), and anthranilate synthase (*trpE/G*), as well as prephenate dehydratase (*pheA2*) and prephenate dehydrogenase (*tyrA2*), which are involved with tryptophan and aromatic amino acid biosynthesis. Additionally, the Env genomes differentially encoded for predicted for lysine 2,3-aminomutase (*kamA*), 4-hydroxy 2-oxovalerate aldolase (*mhpE*), and O-acetylhomoserine (thiol)-lyase (*metY*), all involved in amino acid metabolism. Interestingly, the Fisher's exact tests revealed the Env and HA clades also differ from each other in arginine and proline metabolism by favouring different subsets of genes: the HA genomes were enriched for arginine decarboxylase (*adiA*) and diamine N-acetyltransferase (*speG*), and Env genomes were enriched for cytosine/creatinine deaminase (*codA*). This enrichment in the Env genomes likely suggests there is a limited external source compared to that of the HA environment and as such, biosynthesis of essential amino acids is required.

The Env genomes also appeared to be significantly enriched for genes assigned to KO involved with membrane transport, specifically glycine betaine/proline (*proV/W/X*), molybdate (*wtpA/B/C*), tungstate (*tupA/B/C*), and peptide/nickel (*ABC.PE.A/A1*) transport, as well a drug/metabolite transporter (*TC.DME*) and arsenite transporter *arsB*. Comparatively, the HA genomes were significantly enriched for zinc transport system proteins (*znuB/C*), two uracil permease genes (*pbuG* and *uraA*), and the metal-tetracycline-proton antiporter gene *tetB*. These differences in transport proteins reflect the difference requirements for the HA and Env species, along variable available nutrient between both environments.

Several genes associated with energy metabolism and carbohydrate metabolism were enriched in the Env genomes. Methanogen homocitrate synthase (*aksA*), methanogen homoaconitase large subunit (*aksE/D*), and methanogen homoisocitrate dehydrogenase (*aksF*) were all significantly enriched in

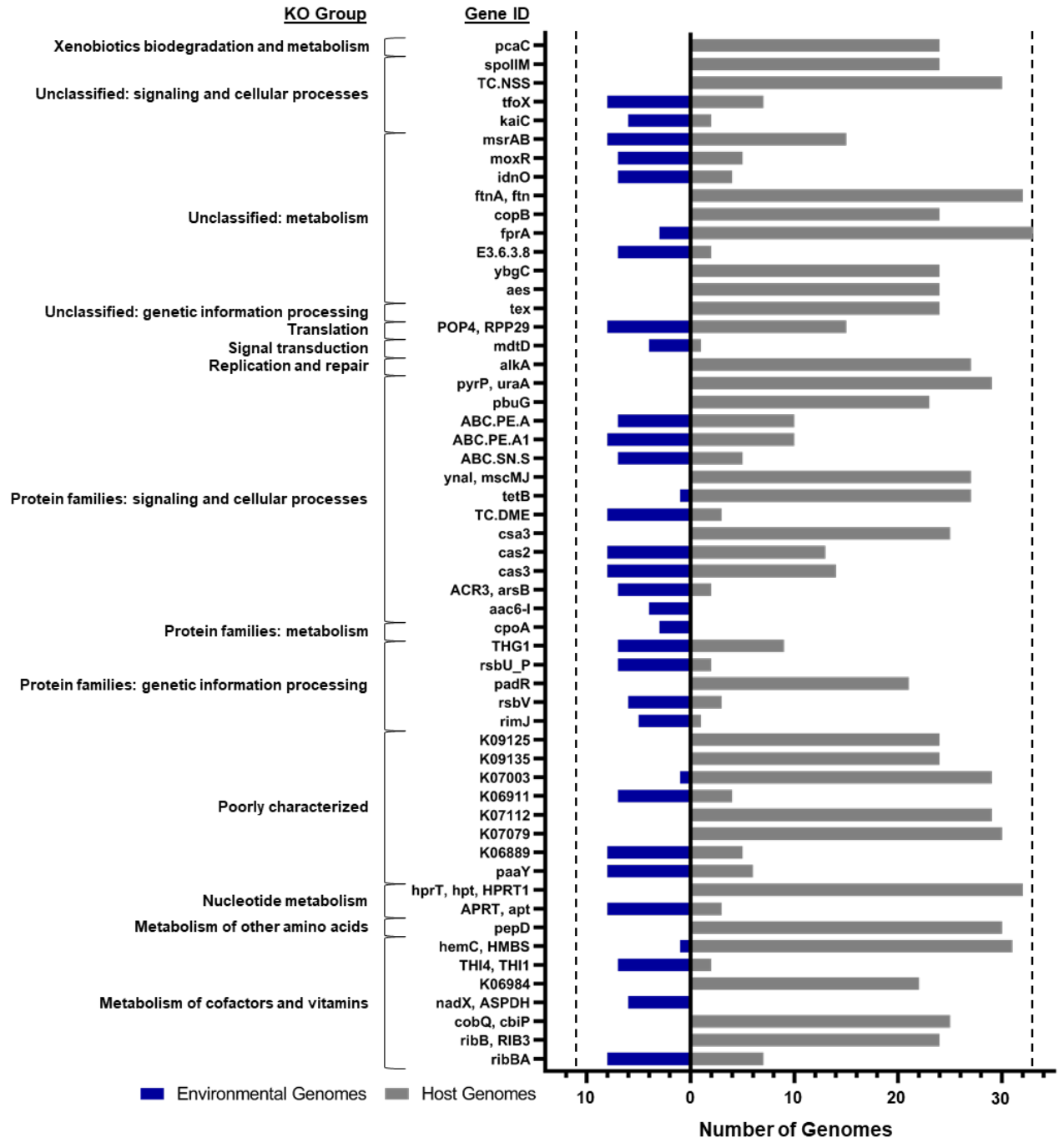


Figure 4.20.A. Genes annotated with KEGG Orthologs enriched in environmental (Env) and host-associated (HA) *Methanocorpusculum* genomes. KO annotation and statistical analysis was performed using the 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). Only KOs with corrected p values of <0.05 were retained and considered significant, as determined by Fisher's Exact Test. 33 HA and 11 Env *Methanocorpusculum* genomes were included in the analysis. See Table 6.18 for the complete list of KOs.

the Env genome, allowing for the biosynthesis of coenzyme B. Acetyl-CoA decarbonylase/synthase (*acsC*) was also significantly enriched in Env genomes and involved in prokaryotic carbon fixation. Additionally, the Env genomes contained a NADP-dependent alcohol dehydrogenase

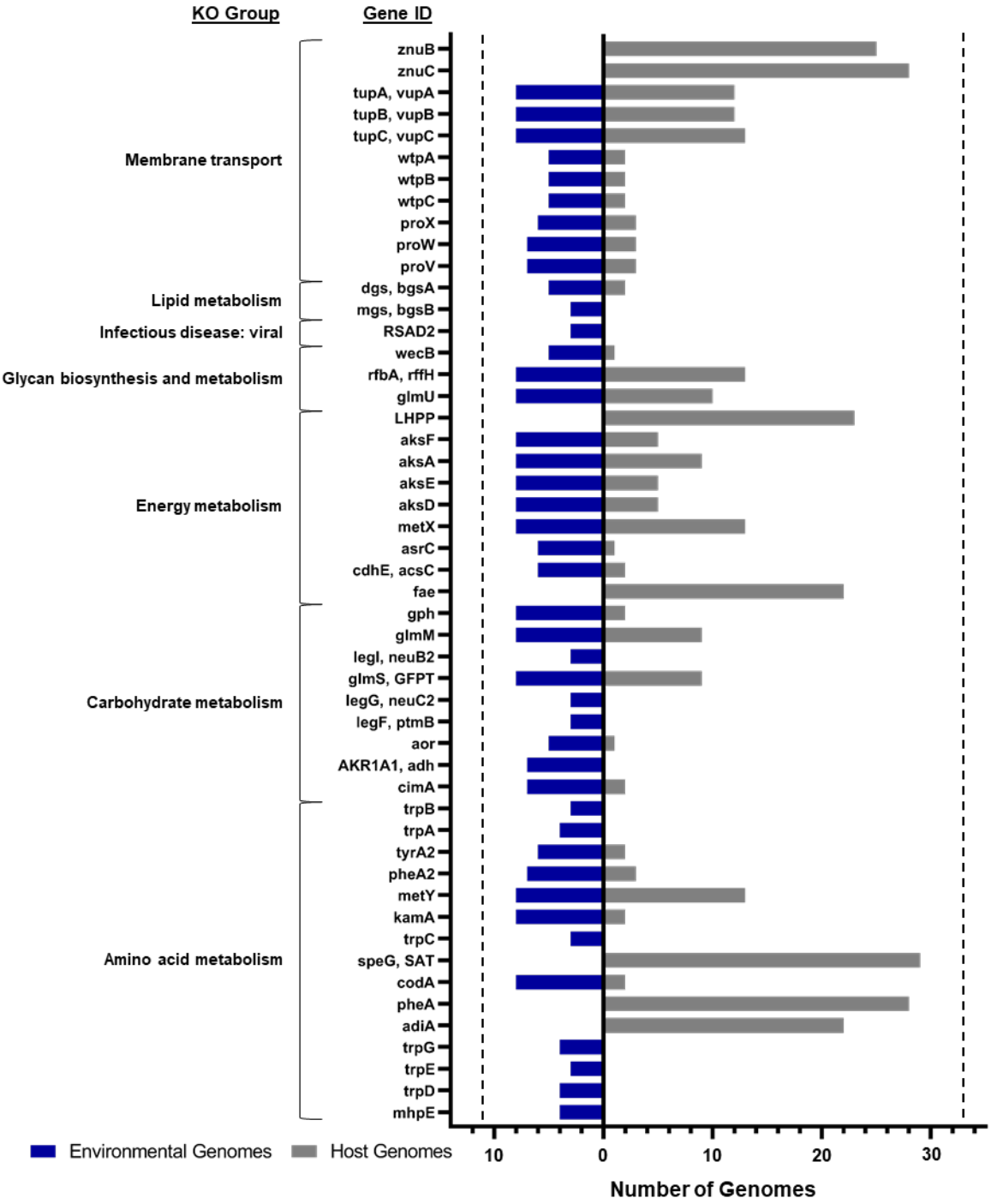


Figure 4.20.B. Genes annotated with KEGG Orthologs enriched in environmental (Env) and host-associated (HA) *Methanocorpusculum* genomes. KO annotation and statistical analysis was performed using 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). Only KOs with corrected p values of <0.05 were retained and considered significant, as determined by Fisher's Exact Test. 33 HA and 11 Env *Methanocorpusculum* genomes were included in the analysis. See Table 6.18 for the complete list of KOs.

(NADP+)(AKR1A1, *adh*) that was not present in the HA genomes and allows for CO_2-dependant methanogenesis with short chain alcohols (Gilmore et al., 2017)(as mentioned in Section 4.3.4). Comparatively, 5,6,7,8-tetrahydromethanopterin hydro-lyase (*fae*) was significantly enriched in the HA genomes, catalysing the formation of methylene-H_4MPT from H_4MPT and formaldehyde, and is notably absent from all Env genomes. However, all genomes also encoded for *fae-hps,* which has additional functionality in ribose phosphate synthesis (Goenrich et al., 2005).

Similarly, Beta-ribofuranosylaminobenzene 5'-phosphate synthase (K06984) also participates in the biosynthesis of methanopterin and is enriched in the HA genomes. Coenzyme F_{420}:H_2 oxidase (*fprA*) is also significantly enriched and the only gene to be enriched in all HA genomes. All 11 genes associated with amino and nucleotide sugar metabolism, O-antigen nucleotide sugar biosynthesis, glycosyltransferases, and glycerolipid metabolism were significantly enriched in the Env genomes. Interestingly, 4-carboxymuconolactone decarboxylase (*pcaC*) was enriched in the HA genomes and *mhpE* was enriched in the Env genomes, both potentially involved in benzoate degradation.

As shown in Figure 4.21, the analysis of the average gene count within the genomes of each group using the Mann-Whitney U test returned fewer significantly enriched genes. In fact, 61% of genes (45/74) significantly enriched by the Mann-Whitney U test were also significantly enriched by Fisher's Exact analysis. The *trp* subunits A-E and G were significantly enriched by Fisher's Exact test but phosphoribosylanthranilate isomerase (*trpF*) was also significantly enriched by Mann-Whitney U test. Although hydroxymethylbilane synthase (*hemC*) was significantly enriched by both analyses, porphobilinogen synthase (*hemB*) was specifically enriched in the Mann-Whitney analysis. Interestingly, several genes involved in glycan biosynthesis and metabolism were significantly enriched, namely *DPM1*, *galE*, *wbjC*, and *capD*. 8-hydroxy-5-deazaflavin:NADPH oxidoreductase (*fno*) was specifically enriched in the Env genomes and is potentially involved in the oxidation of coenzyme F_{420}. Conversely, HA genomes were enriched for cysteine desulphurase/selenocysteine lyase (*sufS*) involved in selenocompound metabolism. *ABC.FEV.A* suggest the Env genomes were similarly enriched for the transport of iron, along with molybdopterin-synthase adenylyltransferase (*moeB*) and sulphur-carrier protein *moaD* involved in the sulphur relay system.

Collectively, these analyses show the Env genomes are significantly enriched for genes associated with amino acid metabolism and biosynthesis. As such, the Env genomes encode for a greater number of biosynthesis genes, where the HA genomes are auxotrophic for many amino acids. This likely reflects a greater availability of exogenous amino acid in the GIT of the animal hosts. Additionally, the Env genomes were enriched for genes associated with coenzyme B biosynthesis, Acetyl-CoA decarbonylase, and an annotated NADP-dependent alcohol dehydrogenase, suggesting the HA genomes encode for a more restrictive substrate specificity for methanogenesis. The HA and Env

genomes are also differentially enriched for transport genes, likely representing clade-specific adaptations.

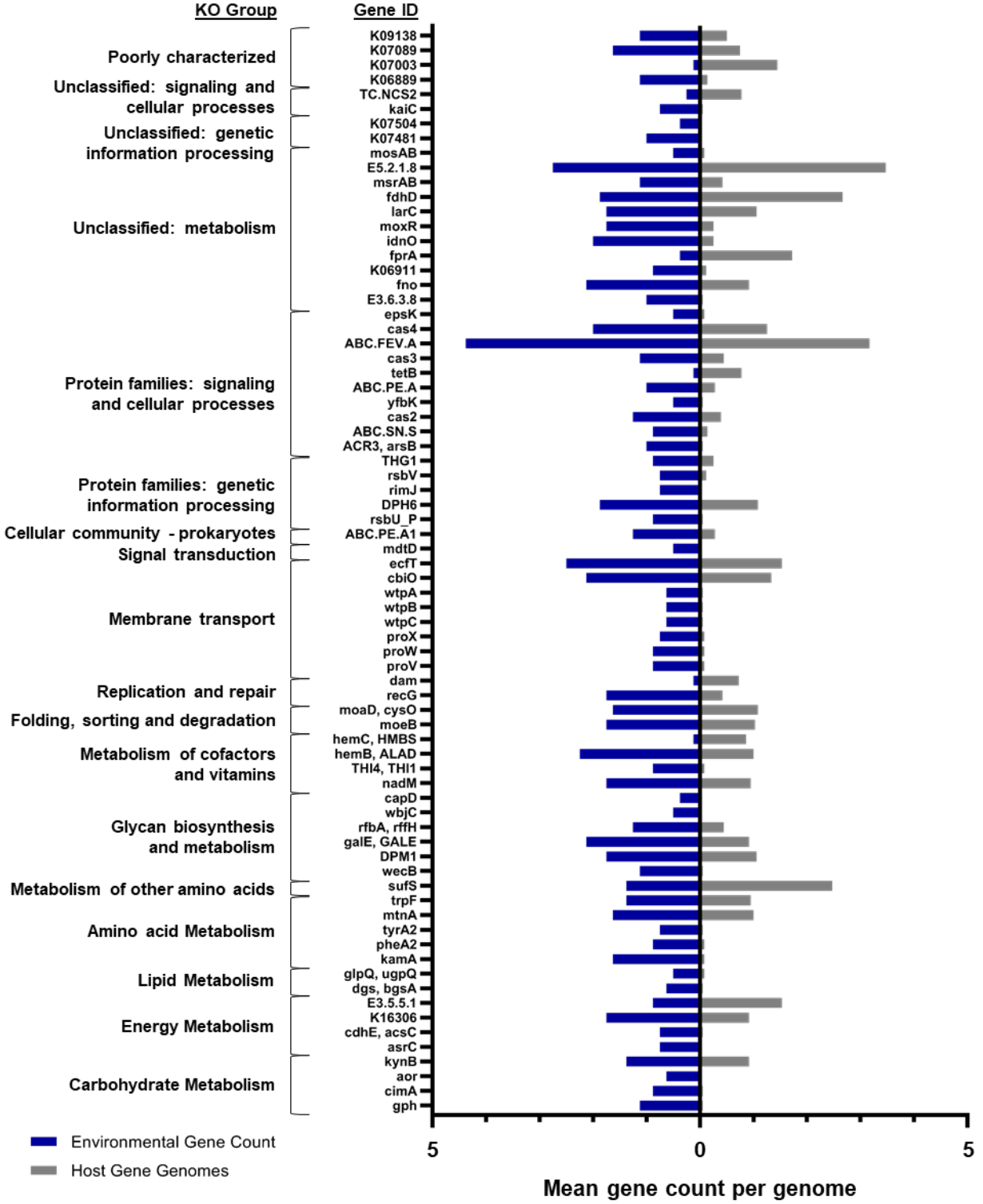


Figure 4.21. KEGG Ortholog gene counts enriched in environmental (Env) and host-associated (HA) *Methanocorpusculum* genomes. KO annotation and statistical analysis was performed using 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). Only KOs with corrected p values of <0.05 were retained and considered significant, as determined by Mann-Whitney U Test. Each value represents the mean gene count per genome for HA (n=33) and Env (n=11) *Methanocorpusculum* groups. See Table 6.19 for the complete list of KOs.

4.3.9 The marsupial-associated *Methanocorpusculum* genomes possess unique genes relative to other host-derived MAGs

Using EnrichM and Fisher's exact test, only 36 KO annotations were differentially represented between the marsupial and other HA genomes, which is considerably less than the 105 differentially enriched KOs between the HA and Env MAGs/genomes. Interestingly, 34/36 KO identified were enriched within the marsupial-associated *Methanocorpusculum*, with only protein ElaA (*elaA*) and ferrous iron transport protein A (*feoA*) enriched in the other HA genomes (Figure 4.22).

Although the predicted propanol-utilizing *adh* of the Env *Methanocorpusculum* is absent from HA genomes, the marsupial genomes were enriched for alcohol dehydrogenase *adh1* (K19954). Interestingly, this gene was found only in the wombat clade, including isolate *Methanocorpusculum* sp. CW153. The Adh1 protein sequence of CW153 showed 56% homology (E= 6e-157) to the predicted phosphonoacetaldehyde reductase of *Natronincola peptidivorans* DSM18979. Interestingly, homology was also shown to the predicted phosphonoacetaldehyde reductase of *Methanosphaera* and *Methanobrevibacter* spp. Similarly, phosphoenolpyruvate phosphomutase (*PPM*; K01841) and phosphonopyruvate decarboxylase (*PPD*; K09459) were also enriched in the marsupial genomes. Additionally, most wombat-associated genomes also contained a methyltransferase gene annotated as telluritemethyltransferase. Blastp of the CW153 protein sequence showed the highest homology to a class I SAM-dependent methyltransferase of *Paenibacillus taiwanensis* (47.20%; E=1e-56) but again showed homology to different *Methanosphaera* spp. Interestingly, these genes are found within a single gene cluster in the *Methanocorpusculum* sp. CW153 genome, flanked by a gene with homology to bacterial transposases (CDS358), suggesting they may have been acquired through horizontal gene transfer. A meso-butanediol dehydrogenase (*budC*) was also enriched in the wombat genomes, which could allow for the utilisation of meso-2,3-butanediol, (S)-Acetoin and/or (S,S)-Butane-2,3-diol for the NADH-dependent production of hydrogen. Additionally, genes annotated as dihydroanticapsinde hydrogenase (*bacC*) may also allow a capacity to produce additional hydrogen.

Benzoyl-CoA reductase subunit A (*badF*) was specifically enriched in all marsupial *Methanocorpusculum* genomes, although annotations were also found in both the macaque-associated MAGs. Though the exact role is unclear due to the absence of subunits B, C, and D, this gene may play a role in the degradation of benzoate compounds in the marsupial hosts, along with the previously mentioned carboxymuconolactone decarboxylase (*pcaC*). Interestingly, Blastp analysis of BadF shows the highest homology (47.81%, E=1e-78) to the 2-hydroxyglutaryl-CoA dehydratase of *Gottschalkia purinilytica*, which participates in amino acid fermentation. Peptide/nickel transport system permease proteins (*ABC.PE.P/P1*), along with peptide/nickel transport system substrate-binding protein *ABC.PE.S*, were also significantly enriched, as was the

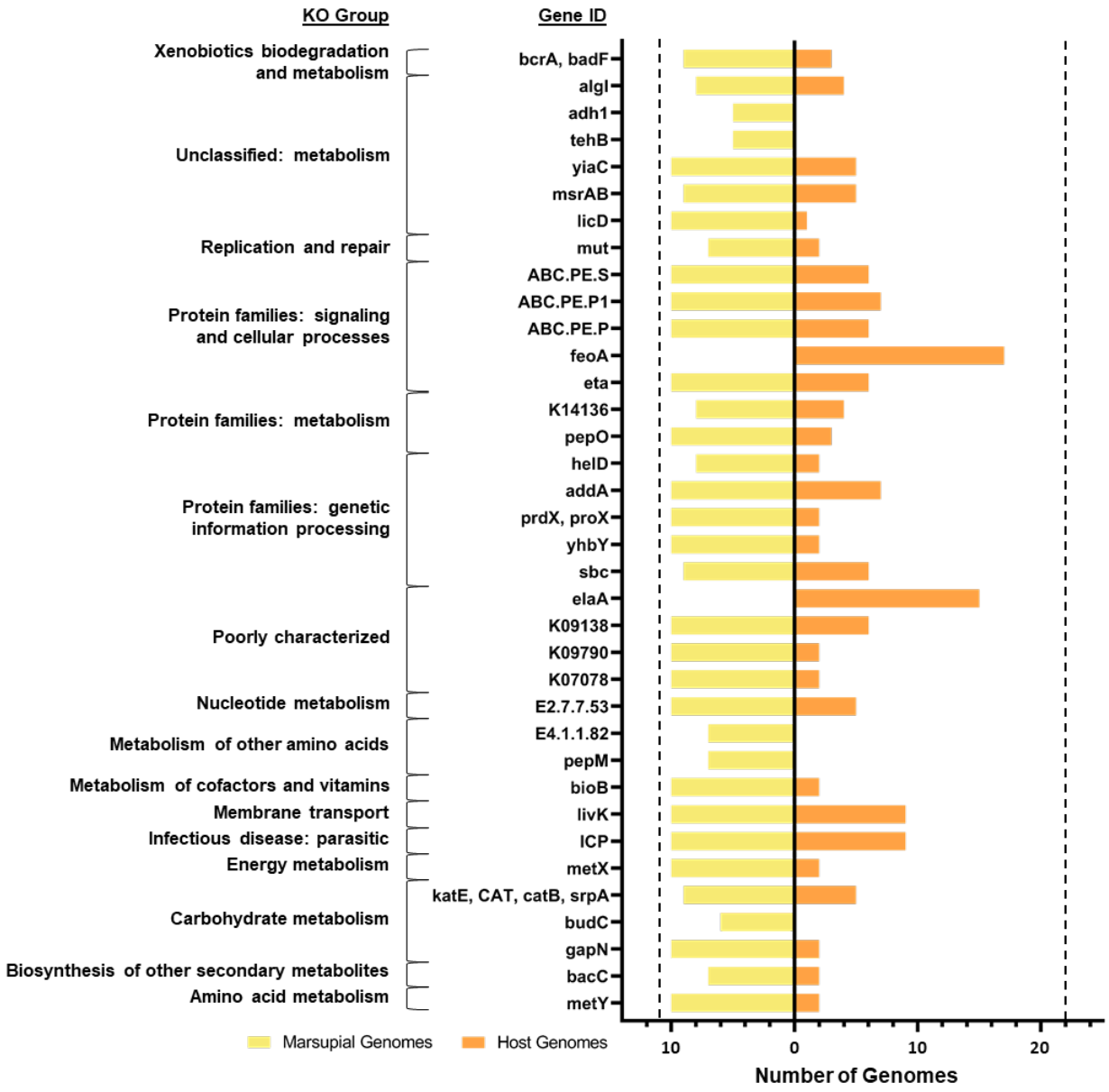


Figure 4.22. Genes annotated with KEGG Ortholog enriched in marsupial- and host-associated *Methanocorpusculum* genomes. KO annotation and statistical analysis was performed using 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). Only KOs with corrected p values of <0.05 were retained and considered significant, as determined by Fisher's Exact Test. 11 marsupial-associated and 22 other host-associated *Methanocorpusculum* genomes were included in the analysis. See Table 6.20 for the complete list of KOs.

capacity for biotin biosynthesis with biotin synthase (*bioB*). Additionally, O-acetyl homoserine(thiol)-lyase (*metY*) and homoserine O-acetyltransferase (*metX*) were also present and involved in methionine and sulphur metabolism. Marsupial genomes were also enriched for alginate O-acetyltransferase complex protein *algI*, decaprenyl-phosphate phosphoribosyl transferase (K14136), and lipopolysaccharide choline phosphotransferase *licD* involved in outer membrane glycosylation and cell wall biosynthesis.

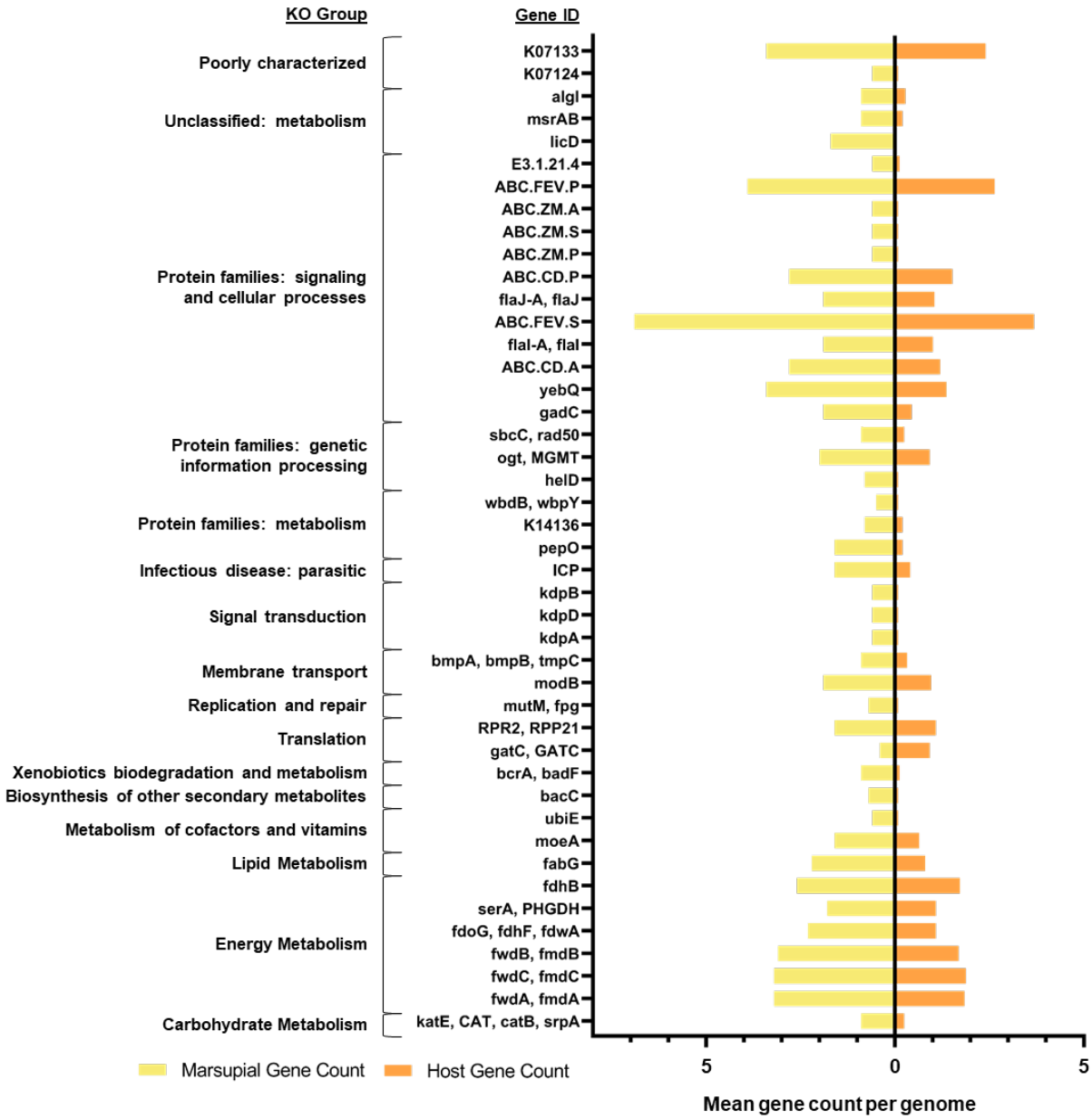


Figure 4.23. KEGG Ortholog gene counts enriched in marsupial- and host-associated *Methanocorpusculum* genomes. KO annotation and statistical analysis was performed using 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). Only KOs with corrected p values of <0.05 were retained and considered significant, as determined by Mann-Whitney U Test. Each value represents the mean gene count per genome for marsupial-associated (n=11) and other HA (n=22) *Methanocorpusculum* groups. See Table 6.21 for the complete list of KOs.

Unlike the HA and Env genome, the Mann-Whitney analysis of the marsupial and other HA genomes provided a larger number of significantly enriched genes than the Fisher's Exact analysis, with 77% being unique genes (Figure 4.23). All genes were significantly enriched in the marsupial genomes, except for aspartyl-tRNA(Asn)/glutamyl-tRNA(Gln) amidotransferase subunit C involved in aminoacyl-tRNA biosynthesis. For the remaining genes, signalling and cellular processes was the

largest category, including zinc/magnesium transport (*ABC.ZM.P/S/A*), iron complex transport (*ABC.FEV.S/P*), putative ABC transport system (*ABC.CD.P*), basic membrane protein A (*bmpA*), and molybdate transport (*modB*). Interestingly, the marsupial genomes were also enriched for flagellar proteins FlaJ and FlaJ. The marsupial genomes contain a greater copy number of formylmethanofuran dehydrogenase subunits (*fwdA/B/C*), along with formate dehydrogenase (*fdhF*), formate dehydrogenase (coenzyme F420)(*fdhB*), and D-3-phosphoglycerate dehydrogenase (*serA*) also involved in methane metabolism. Similarly, molybdopterin molybdotransferase was also enriched and plays an important role in molybdopterin biosynthesis for methanogenesis. K+-transporting ATPase A (*kdpA*), K+-transporting ATPase B (*kdpB*), and sensor histidine kinase *kdpD* were also significantly enriched and involved in potassium transportation.

Despite the phylogenetic similarity of the marsupial *Methanocorpusculum* lineages, each shared differentially enriched genes (Figure 4.24), though the small number of genomes in the marsupial groups limited statistical analyses. As previously mentioned, the wombat-associated genomes were significantly enriched for the *adh1*, meso-butanediol dehydrogenase *budC*, and tellurite methyltransferase *tehB* (Figure 4.24). Proteins associated with uptake and transport appear to be differentially enriched between the strains, with the wombat-associated genome enriched for oligopeptide transport system ATP-binding proteins *oppD/F*, K+-transporting ATPase *kdpA/B*, phosphate transport system substrate-binding protein *pstS*, putative sodium/glutamine symporter *glnT*, tungstate transport system proteins *tupA/B*, and zinc/manganese transport system proteins *ABC.ZM.A/P/S*, along with a tetracycline resistance efflux pump (*tet35*). Comparatively, the mahogany glider-associated genomes were enriched for alanine or glycine:cation symporter *TC.AGCS*, peptide/nickel transport system proteins *ABC.PE.A/A1*, Na+:H+ antiporter *nhaC*, arabinogalactan oligomer/maltooligosaccharide transport system substrate-binding protein *cycB*, and zinc transport system substrate-binding protein *znuA*. These adaptions show the differential availability of available trace elements in the GIT of the different hosts and the requirements for each species.

Nitrogenase molybdenum-cofactor synthesis protein (*nifE*), nitrogenase molybdenum-iron protein (*nifN*), nitrogenase molybdenum-iron protein components (*nifD*, *nifK*), and nitrogen regulatory protein PII components one and two were all enriched in the mahogany glider-associated genomes, suggesting the presence of a more diverse nitrogen assimilation pathway (as described in Section 4.3.4). The presence of acetolactate synthase I/II/III (*ilvB, ilvG, ilvI*), acetolactate synthase I/III (*ilvH*, *ilvN*), threonine dehydratase (*ilvA*), and dihydroxy-acid dehydratase *ilvD* in the mahogany glider genes also allows for the biosynthesis of valine and isoleucine (as described in Section 4.3.4). Despite methanogen homocitrate synthase being enriched in the Env genomes, the marsupial genomes were

the only HA clade to also contain this gene. The mahogany glider genomes were also enriched for

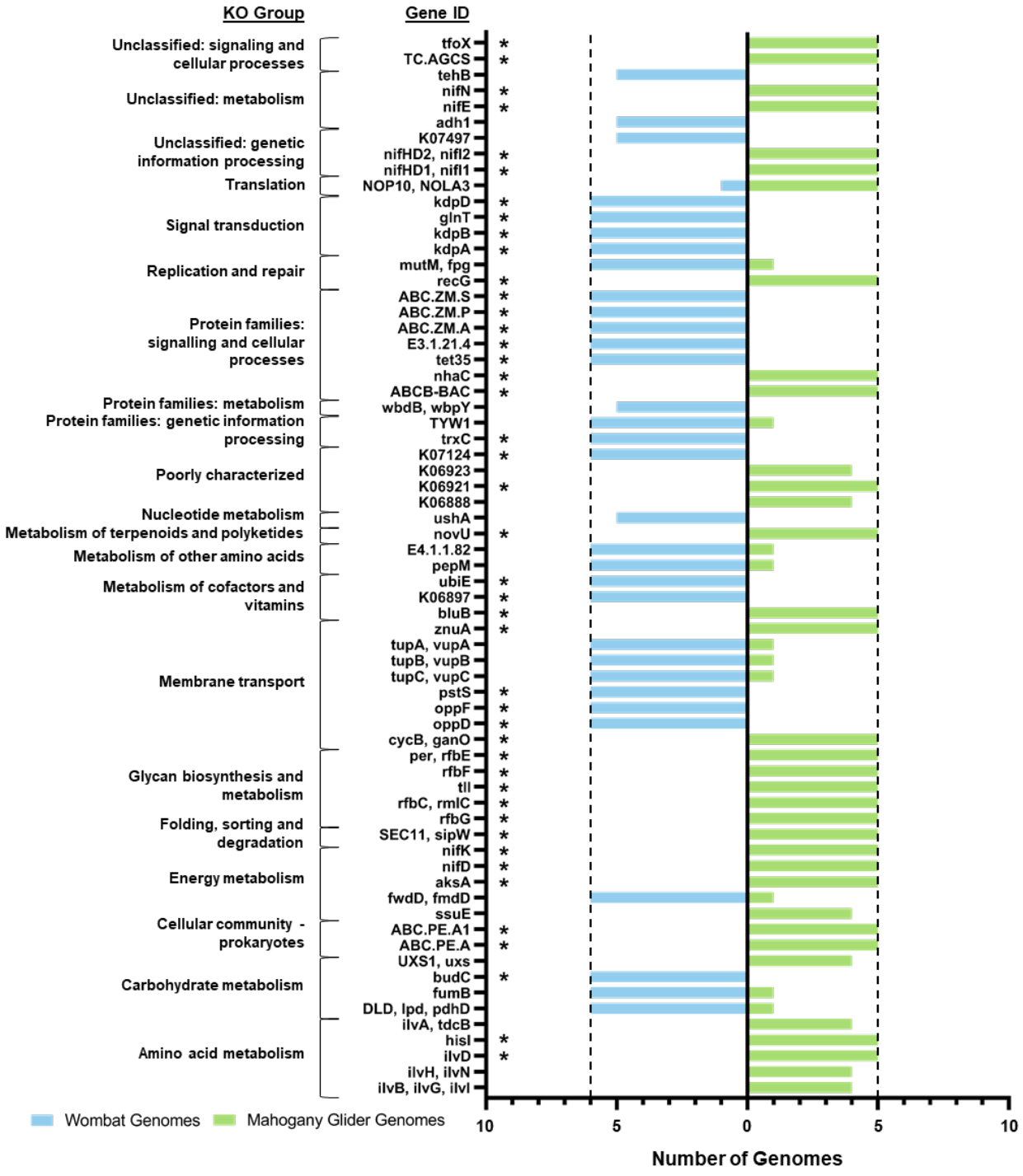


Figure 4.24. Genes annotated with KEGG Orthologs enriched in wombat- and mahogany glider-associated *Methanocorpusculum* genomes. KO annotation and statistical analysis was performed using 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). Genes with p values of <0.05 were retained and those with a corrected p value of <0.05 were considered significant, as determined by Fisher's Exact Test. Genes with statistically significant enrichment are shown by *. Six wombat-associated and five mahogany glider-associated *Methanocorpusculum* genomes were included in the analysis. See Table 6.22 for the complete list of KOs.

genes annotated as CDP-glucose 4,6-dehydratase (*rfbG*; K01709), UDP-glucuronate decarboxylase (*uxs*; K08678), and glucose-1-phosphate cytidylyltransferase (*rfbF*; K00978), predicted to participate in starch and sucrose metabolism, and O-antigen nucleotide sugar biosynthesis. Conversely, the wombat genomes were enriched for the *fumB* component of fumarate hydratase predicted for the conversion of (S)-malate to fumarate (Katayama et al., 2019).

Interestingly, the wombat-associated clade contained subunit D of formylmethanofuran dehydrogenase. This clade also contains 7,8-dihydropterin-6-yl-methyl-4-(beta-D-ribofuranosyl)aminobenzene 5'-phosphate synthase, which produces the precursor for 5,6,7,8-tetrahydromethanopterin (H_4MPT) biosynthesis (Xu et al., 1999). Although enriched in the marsupial genomes, phosphoenolpyruvate phosphomutase and phosphonopyruvate decarboxylase are present in all wombat genomes and only a single mahogany glider genome. The wombat genomes also appeared to contain demethylmenaquinone methyltransferase (*ubiE*) and dihydrolipoamide dehydrogenase (*pdhD*). Mahogany glider genomes were enriched for both FMN reductase (*ssuE*) and

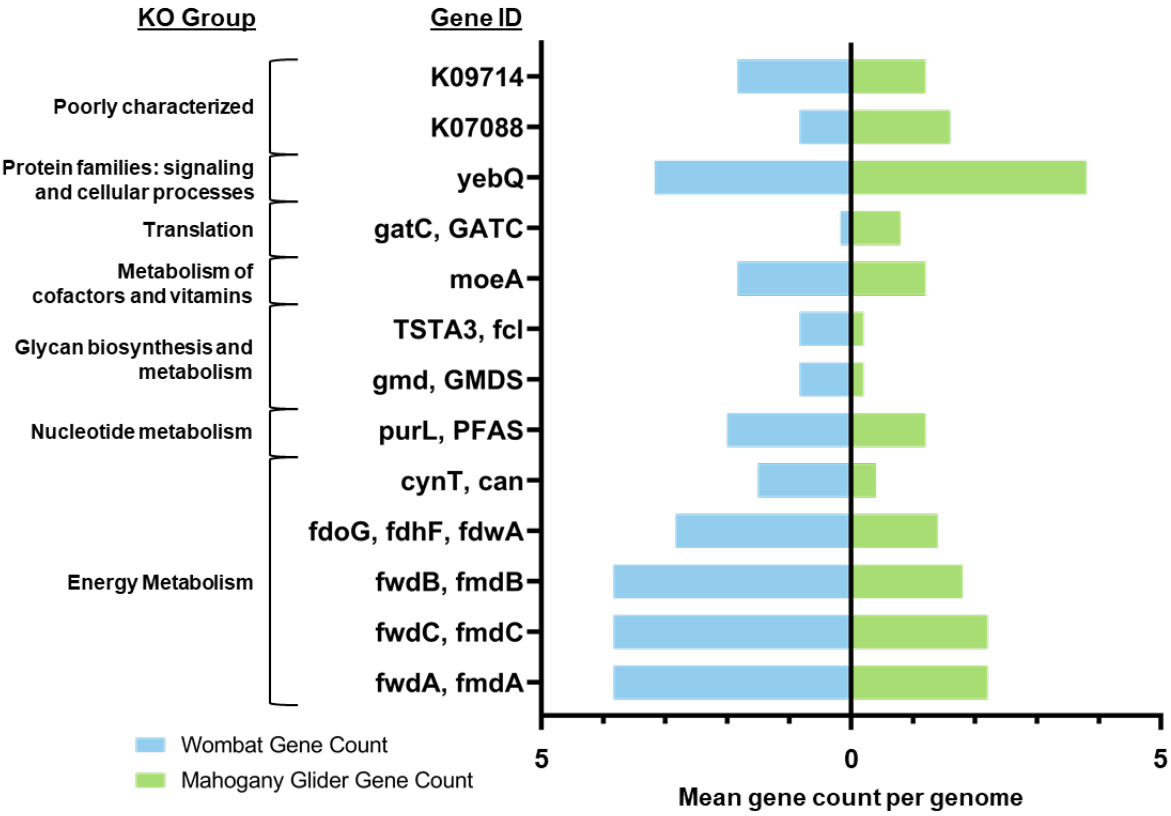


Figure 4.25. KEGG Ortholog gene counts enriched in wombat- and mahogany glider-associated *Methanocorpusculum* genomes. KO annotation and statistical analysis was performed using 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). Only KOs with p values of <0.05 were retained and considered significant, as determined by Mann-Whitney U Test. Each value represents the mean gene count per genome for wombat-associated (n=6) and mahogany glider-associated (n=5) *Methanocorpusculum* groups. See Table 6.23 for the complete list of KOs.

5,6-dimethylbenzimidazole synthase (*bluB*). Perosamine synthetase (*per*, *rfbE*), alpha-1,3-rhamnosyltransferase (*wbdB*, *wbpY*) and dTDP-4-dehydrorhamnose 3,5-epimerase (*rfbC*, *rmlC*) were similarly enriched.

Thirteen genes were differentially enriched between the two groups by the Mann-Whitney U test, although none were statistically significant after P value correction (Figure 4.25). Similar to the marsupial and HA analyses, formylmethanofuran dehydrogenase (*fwdA/B/C*) and formate dehydrogenase (*fdhF*) were enriched in the wombat genomes, along with molybdopterin molybdotransferase (*moeA*). Interestingly, carbonic anhydrase (*can*) was enriched in the wombat genomes and likely also involved in nitrogen metabolism. Similarly, O-antigen nucleotide sugar biosynthesis-associated genes GDPmannose 4,6-dehydratase (*gmb*) and GDP-L-fucose synthase (*TSTA3*) were also enriched in the wombat genomes. Only three genes were enriched in the mahogany glider genomes: aspartyl-tRNA(Asn)/glutamyl-tRNA(Gln) amidotransferase subunit C, multidrug resistance protein *yebQ* and uncharacterised protein K07088.

Collectively, key differences were observed in glycosylation, membrane transport, O-antigen biosynthesis, amino acid metabolism and the metabolism of various secondary metabolites. This differential enrichment of genes between the marsupial-associated and other host-associated *Methanocorpusculum* show the unique genetic adaptations of the different HA clades to their host enrichment. Additionally, differences in genes between the wombat and mahogany glider clades provides further evidence that they represent distinct marsupial-associated species.

4.3.10 The marsupial-associated *Methanocorpusculum* genomes encode for unique carbohydrate active enzymes

There was also a significant variation of predicted cazymes between *Methanocorpusculum* lineages (Figure 4.26). Interestingly, a gene annotated as auxiliary activity (AA) 6 was encoded by all genomes, however, the gene counts were significantly greater in the marsupial-associated lineages (P=0.0026), with 80% (4/5) mahogany glider genomes containing three annotated AA6 genes. Two carbohydrate esterase (CE) categories were detected across the genomes; CE1 and CE10. The CE genes were detected in three clades of HA *Methanocorpusculum*, with CE1 was detected in the chicken and mahogany glider clades and the two marsupial clades containing CE10. The CE1 genes of *Methanocorpusculum* sp. MG showed the greatest similarity to an alpha/beta fold hydrolase protein of *Methanosarcinaceae* at 32.49% (1e-31). The prokka-based annotation suggested that the gene was an aminoacrylate hydrolase (EC:3.5.1.-; K09023) involved in pyrimidine metabolism. The CE10 family was significantly enriched in the marsupial-associated clades (Fisher P=1.01E-05, Mann-Whitney P=4.67E-06), with one additional chicken-derived MAG (Figure 4.26). Like the wombat

genomes, the macaque-derived MAGs also contained CE10 annotated genes. Both proteins of *Methanocorpusculum* sp. MG and CW153 show no significant similarity to any archaeal species, suggesting these annotations are unique to these lineages.

Four glycoside hydrolase (GH) families were detected, each represented by a single gene annotation. Mahogany glider genome S0463_NS37_NS81_mahogany_glider_3 contained two of these annotations, GH105 and GH16. *Methanocorpusculum* sp. GPch4 contained a GH133, likely with amylo-α-1,6-glucosidase-like (EC 3.2.1.33) activity. Additionally, genome

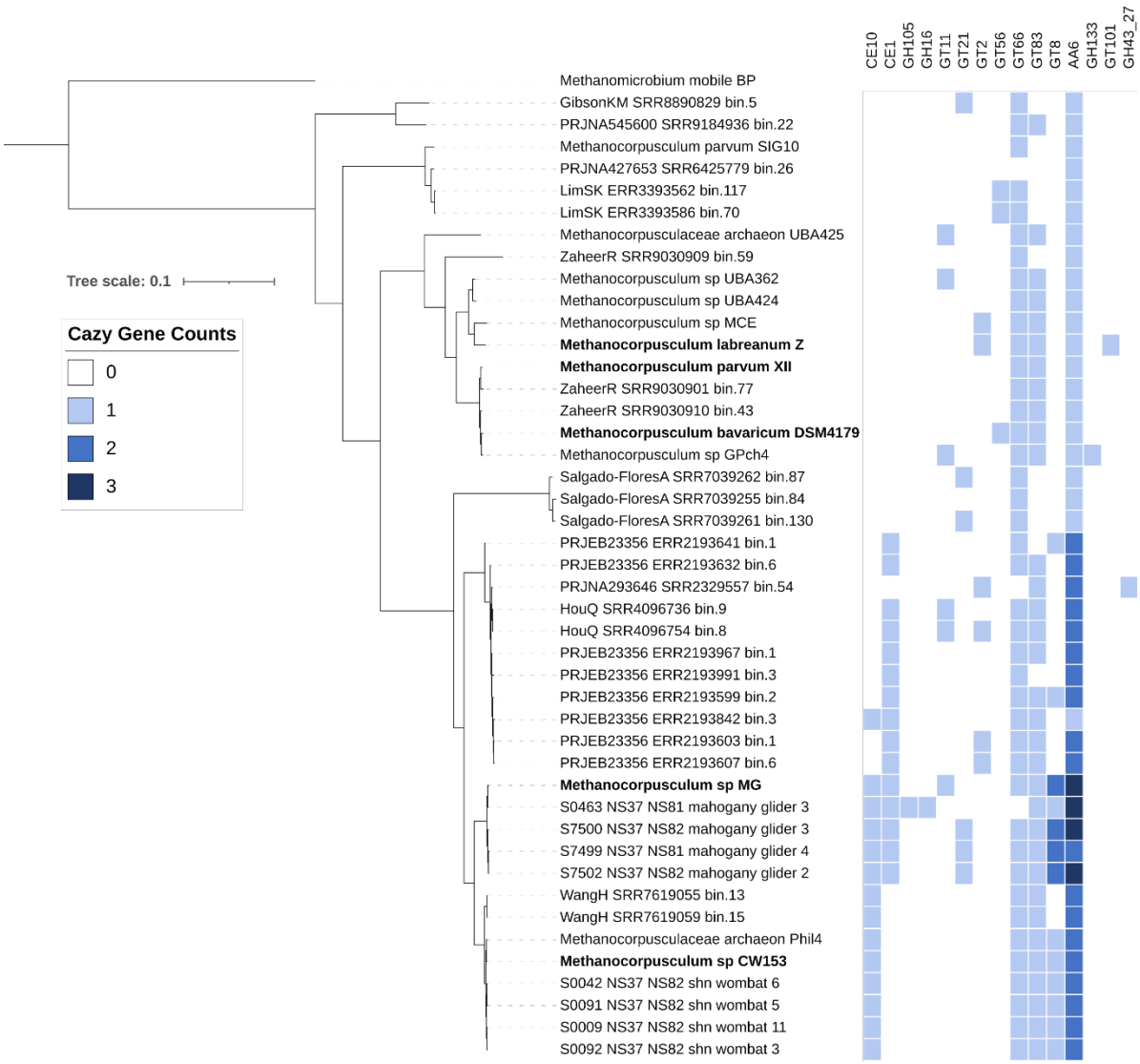


Figure 4.26. Phylogenetic tree of *Methanocorpusculum* showing the distribution of annotated carbohydrate active enzymes. The phylogenetic tree was produced using GTDB-tk (v.1.3.0), FastTree (v2.1.10) and visualisation by iToL (https://itol.embl.de/), with *Methanomicrobium mobile* BP used as the outgroup. Cazymes were annotated with EnrichM (v0.5.9). The marsupial-associated lineages were enriched for a specific subset of AA, CE and GT annotated genes.

PRJNA293646_SRR2329557_bin.54 contained a protein annotated as GH43_27 with predicted arabinose, arabinofurinase, and xylosidase activity. No cultured isolate contained annotated GH enzymes; thus, experimental confirmation of these findings may prove challenging. However, these initial analyses suggest that specific genomes within *Methanocorpusculum* may contain the potential to hydrolyse different carbohydrates.

Glycosyltransferases (GTs) represent the largest variety of annotated cazymes with eight unique categories. GT66 and GT83 appear to be conserved over most genomes, with the latter only absent from ptarmigan- and goat-derived genomes. Interestingly, the GT8 annotations were significantly enriched in the marsupial lineages, although the macaque-derived genomes did not contain any genes with GT8 annotation. Additionally, the mahogany glider genomes showed a greater copy number of GT8, associated with lipopolysaccharide galactosyl/glucosyl/glucuronyltransferases (Lombard et al., 2014). GT105, GT16, GT11, GT21, GT2, GT56, and GT101 were also detected in low counts across the genomes (Figure 4.26). Interestingly, some of the genomes from cultured isolates contain unique cazyme annotations. *Methanocorpusculum* sp. MG contains a GT11 annotated gene only sparsely observed across other lineages. Blast analysis of this protein shows the highest similarity (35.22%) to alpha-1,2-fucosyltransferase of *Methanocorpusculum* sp. GPch4 (E=6e-55), which was also annotated as GT11. *M. bavaricum* contains a GT56 and *M. labreanum* contains GT2 and GT101 annotated genes, with the latter two only present in the *M. labreanum* genome. These analyses show a diverse glycosylation potential by the different genomes of *Methanocorpusculum*, with a specific enrichment of unique cazymes encoded in the marsupial-associated *Methanocorpusculum.*

4.3.11 Substrate utilisation profiles of marsupial-associated *Methanocorpusculum* isolates

Previously analysis of *Methanocorpusculum* has shown *M. parvum* and *M. bavaricum* to utilise short-chain alcohols in CO_2-dependent methanogenesis, with *M. parvum* able to utilise 2-propanol and 2-butanol (Zellner et al., 1989). Analysis of the *adh* genes also predicted utilisation of cyclopentanol, 2,3-butanediol, ethanol, and 1-propanol for *M. parvum,* with the *adh* of *M. bavaricum* predicted to utilise cyclopentanol and 2,3-butanediol (Bleicher et al., 1989). Interestingly, my analysis by EnrichM showed only the HQ MAGs and isolate genomes from environmental sources (i.e. *M. parvum, M. bavaricum,* and *M. labreanum*) to be enriched for homologs of the gene predicted to support propanol utilisation by *M. parvum* (Section 4.3.8)(Bleicher & Winter, 1991).

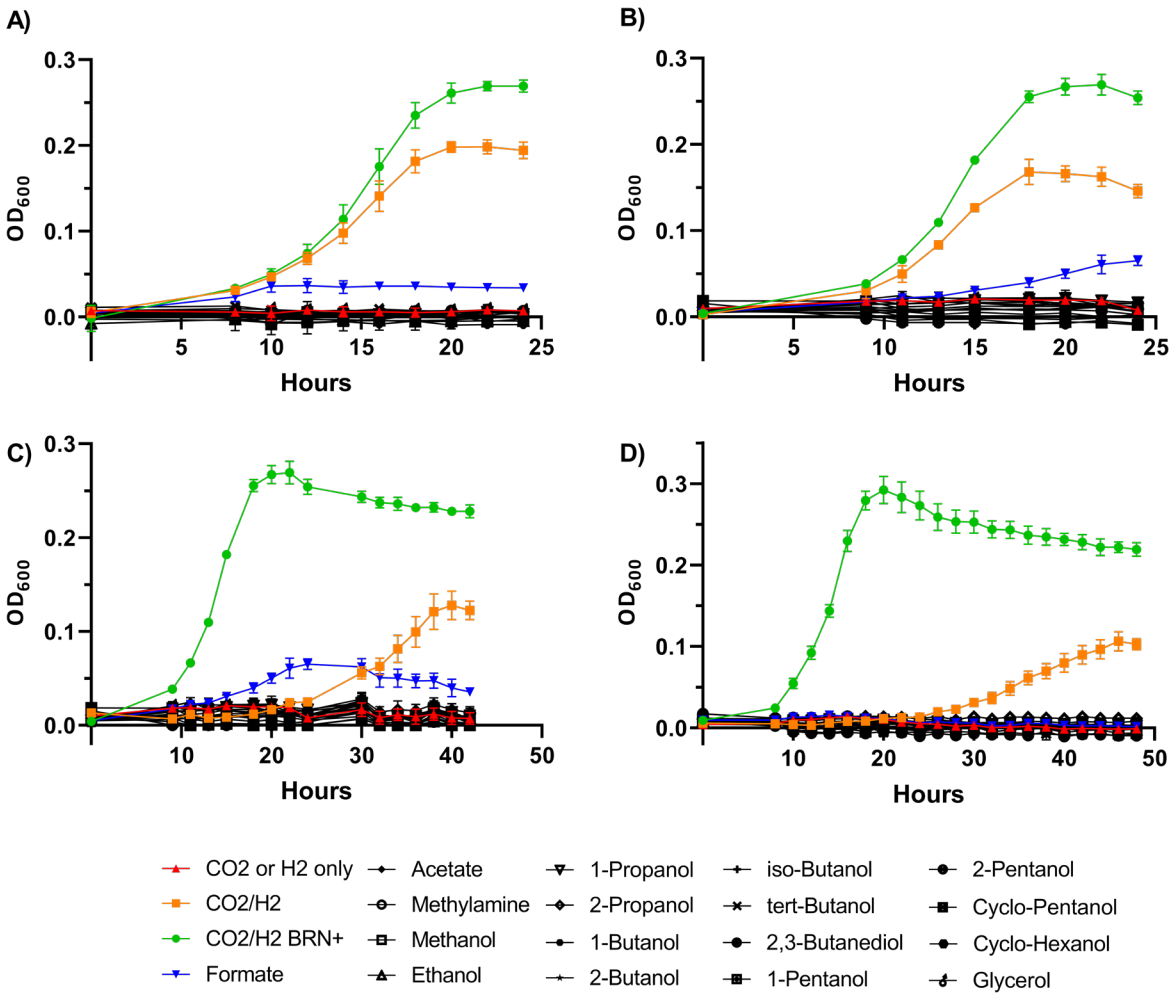


Figure 4.27. Primary substrate utilisation of *Methanocorpusculum* sp. CW153 and MG *in vitro*. Each strain was grown in broth BRN-RF10 (sodium formate and sodium acetate removed) culture at 37ºC with 1% (v/v) supplementation of test substrates. **A)** and **B)** represent *Methanocorpusculum* sp. CW153 with a headspace of CO_2 and H_2, respectively. **C)** and **D)** represent *Methanocorpusculum* sp. MG with a headspace of CO_2 and H_2, respectively. CO_2 and H_2 alone were used for negative controls. CO_2/H_2 was used for the positive control. CO_2/H_2 with 'BRN+' (sodium formate and sodium acetate added) was used to show growth in basal BRN-RF10 medium.

Based on these differences such, I chose to culture *Methanocorpusculum* sp. MG and CW153 with various short-chain alcohols as primary substrates for methanogenesis. As Figure 4.27 shows, both species could effectively utilise CO_2 and H_2 for growth, although the removal of sodium acetate and sodium formate from the medium significantly reduced the maximum yield of both strains. Some of this reduction in yield can be attributed to the lack of sodium formate as the cultures supplemented with formate showed significant growth, although these cultures produced a lower maximum yield compared to CO_2 and H_2. Furthermore, neither *Methanocorpusculum* sp. MG nor CW153 showed any significant growth with short-chain alcohols for methanogenesis, as shown in Figures 4.27, even

with prolonged incubation of the cultures (500 h). *Methanocorpusculum* sp. MG grown under a headspace of CO_2 and with 1-butanol supplementation did produce a small increase in yield (OD_{600}= 0.022) after ~200 h of incubation (Figure 4.28). Such findings suggest that with the culture conditions tested here, both strains CW153 and MG were effectively incapable of growth when short-chain alcohols were provided as a potential carbon source, or as an alternative reductant to H_2 gas.

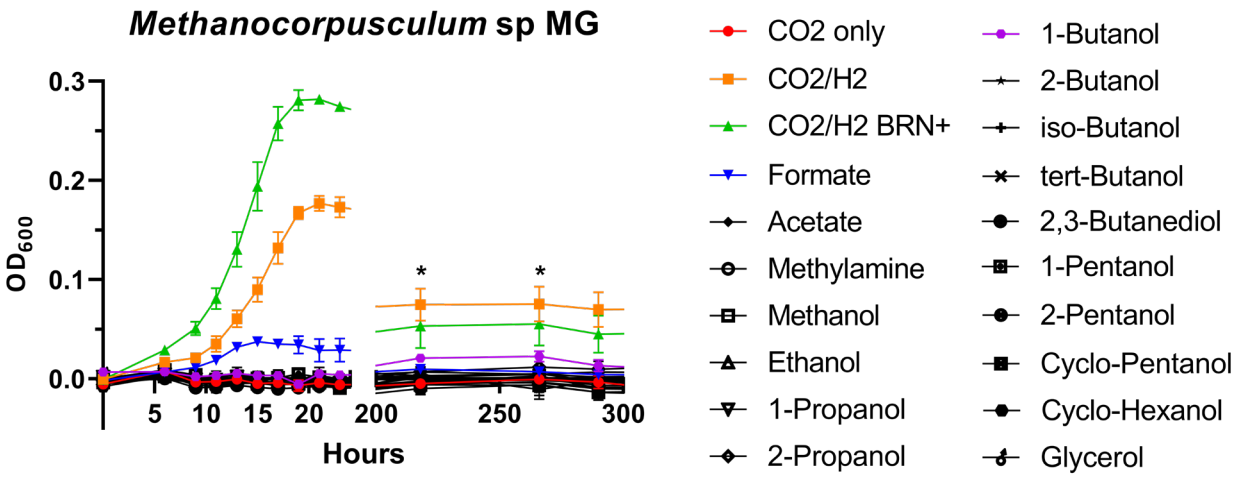


Figure 4.28. Substrate analysis of *Methanocorpusculum* sp. MG in the presence of CO_2. *Methanocorpusculum* sp. MG was grown with substrates of interest and a headspace of CO_2 (same study to figure 4.26C). No additional substrates showed any significant growth, except for 1-butanol after 200 h which showed a statistically significant increase (max mean OD_{600}= 0.022)(P = 0.0032) in yield by OD_{600} compared to CO_2 only.

4.4 Discussion

Previous studies with kangaroos and wallabies suggest that these animals produce less methane from digestible energy compared to ruminant livestock, while maintaining a similar diversity of methanogens (Evans et al., 2009; Von Engelhardt et al., 1978; Zellner et al., 1989). Although previous studies have attempted to characterise the microbiome of other marsupial species, Bacteria-Domain-specific 16S rRNA gene primers have been used, which might have limited the holistic characterisation of the marsupial-associated gut archaeal communities (Barker et al., 2013; Burke et al., 2018; Cheng et al., 2015; Eisenhofer et al., 2021). However, "universal" 16S rRNA gene primers have also failed to identify any archaeal species from kangaroo and koala faecal samples (Brice et al., 2019; Dahlhausen et al., 2018; Gulino et al., 2013). Here, I have used "universal" 16S rRNA gene primers with DNA extracted from faecal samples, and successfully identified the presence of candidate novel lineage(s) of host-associated *Methanocorpusculum* and/or *Methanobrevibacter* lineages across 23 marsupial species. In contrast to previous studies, my 16S rRNA-based surveys did not identify *Methanosphaera* or *Methanomassiliicoccales* in the faecal samples examined, although both were detected previously in samples of foregut digesta from macropodids (Klieve et al., 2012). Furthermore, the *Methanocorpusculum*-affiliated reads were separated into two OTUs, with each OTU differentially distributed across the marsupial specimens examined. Contrary to the amplicon-based sequencing, the MGS datasets from most of the marsupial specimens produced no detectable methanogen signal. This could be attributed to a low abundance of the methanogen populations, which require an increased depth of sequencing to accurately capture, except in the case of wombats, koala, and brushtail possum. Overall, the shotgun metagenome datasets from the same specimens showed a reduced prevalence of methanogens compared to the amplicon profiles, with none of the kangaroo or wallaby faecal samples possessing a detectable methanogen population via MGS. It is worth noting that only one koala faecal sample had detectable methanogens by MGS, which was the same faecal sample that showed a high percentage of *Methanobrevibacter* by amplicon sequencing. In those samples with a detectable methanogen community by MGS, the diversity revealed was greater than the 16S rRNA amplicon profiles, including the presence of *Methanobacteriaceae, Methanocorpusculum*, *Methanosarcina* and *Methanomassiliicoccaceae*. These results reflect the advantages and disadvantages associated with 16S rRNA gene amplicon-based and MGS approaches. The results are consistent with PCR amplicon libraries offering a more sensitive but less informative assessment of microbial taxa, while MGS data can provide a more detailed resolution of microbial diversity, its limits of detection are relatively high and crucially affected by sequencing depth. Additionally, the use of archaeal-specific 16S primers, or other conserved methanogen genes such as *McrA*, would also improve the detection of the diverse archaeal community that is not accurately captured through the use of universal 16S amplification.

Interestingly, the red-necked wallaby and red kangaroo faecal cultures produced a greater percentage of methane (3.052±0.089%, 2.757±0.494%) compared to the common and Southern hairy-nosed wombat (2.037±2.880%, 2.137±0.231%). However, the wombat samples contained a greater abundance of methanogens, specifically *Methanocorpusculum*_1 and *Methanocorpusculum*_2, as determined by 16S rRNA sequencing. This may be attributed to these kangaroo samples containing a higher abundance of *Methanobrevibacter*, which has been associated with a high methane producing phenotype in bovine (Martínez-Álvaro et al., 2020). Alternatively, the culture conditions used to propagate the faecal samples may better support the growth of the kangaroo-associated methanogens compared to the wombat, and thus selectively increase the methane production. Despite the production of measurable amounts of hydrogen in anaerobic culture, most (but not all) koala faecal samples only produced trace concentrations of methane, although some samples were found to contain detectable populations of *Methanobrevibacter, Methanocorpusculum*, and *Methanomassiliicoccales*. Additionally, I was unable to recover methanogen positive enrichments from the koala specimens I examined. There are multiple possible reasons for these findings, including the impacts of sample preservation and storage on methanogen viability. Variations in the choice of *Eucalyptus* foliage by different koalas might also directly suppress methanogen growth and/or methane formation (Cork et al., 1983; Eschler et al., 2000). Furthermore, the comparative metagenomics study by Shiffman et al. (2017) suggest that the koala microbiome favoured the presence and metabolic activity of sulphate-reducing bacteria rather than methanogens and methanogenesis.

While my efforts using the koala faecal samples were unsuccessful, I did produce methane-positive enrichments leading to the isolation of two novel *Methanocorpusculum*-affiliated methanogens, each from a mahogany glider (MG) and common wombat (CW153) faecal sample. To my knowledge, these are the first cultured isolates of *Methanocorpusculum* to be recovered from animal hosts. The comparative MGS studies by Shiffman et al. (2017) did identify *Methanocorpusculum* spp. in the faecal sample of a Southern hairy-nosed wombat at a relative abundance of 2.14% compared to the koala faecal sample at 0.11%. This MGS dataset from the wombat sample also produced a MAG affiliated with the genus *Methanocorpusculum*, with isolate *Methanocorpusculum* sp. CW153 representing a cultured representative of the same species. Comparatively, *Methanocorpusculum* sp. MG clusters as a separate but closely related clade, representing an additional novel species. Interestingly, *Methanocorpusculum* sp. MG and CW153 are both representatives of the OTU *Methanocorpusculum*_2, as described in Section 4.3.2, with 100% sequence identity to *Methanocorpusculum*_2 and 99.6% to *Methanocorpusculum*_1. Not all recovered MAGs contained

16S rRNA genes and so could not be separated into a *Methanocorpusculum* OTU group, however, MAGs from chicken, mahogany and wombat clades were all representatives of *Methanocorpusculum*_2 (100% similarity). MAGs derived from humans and cows were also closely related to *Methanocorpusculum*_2, albeit with only 98.8% sequence similarity. There was no perfect representative of the *Methanocorpusculum*_1 OTU, although the Env genomes were the closest with 99.2% similarity.

The previously isolated strains of *Methanocorpusculum* spp. were only derived from environmental sources and were remarkable for their ability to utilise 2-propanol and 2-butanol as a source of hydrogen, in addition to typical hydrogenotrophic methanogenesis with CO_2 and H_2 (Bleicher et al., 1989; Zellner et al., 1989). Further characterisation of these *Methanocorpusculum* and other closely related *Methanomicrobiales* isolates showed their potential to also utilise cyclopentanol, 2,3-butanediol, ethanol, and 1-propanol (Bleicher & Winter, 1991). In a similar vein, *Methanosphaera* sp. WGK6 isolated from the Western grey kangaroo can utilise ethanol as a substitute for hydrogen for methanol-dependent methanogenesis (Hoedt et al., 2016). For these reasons, I tested whether *Methanocorpusculum* sp. CW153 and/or MG show a similar substrate utilisation profile to these archaea. However, my results suggest that, under the growth conditions tested, neither strain can utilise alcohols as primary substrates; and are restricted to using hydrogen, carbon dioxide and/or formate for methanogenesis and growth. In that context, neither isolate genome, nor the HQ MAGs of host-origin were found to encode a homolog of the *M. parvum adh* believed to catalyse the use of short-chain alcohols as a carbon source (Bleicher & Winter, 1991). This may suggest the availability of short-chain alcohols is not apparent in these host-associated environments and as such, is not a beneficial adaptation for these methanogens. However, at least some of the HA MAGs and isolates genomes possess gene homologs of other dehydrogenases implicated in the reduction of short chain alcohols, including *Methanocorpusculum* sp. CW153 and MG.

My findings confirm that lineages affiliated with the genus *Methanocorpusculum* are bona fide residents of the digestive tracts of a broad range of vertebrate animals from terrestrial and marine environments. Indeed, I have successfully recovered *Methanocorpusculum* MAGs from publicly available MGS datasets, expanding the number of available genomes by 134 (83%) and providing genomes recovered from 13 host species. As shown in Figures 4.11-4.13, phylogenetic and AAI analyses showed at least 18 host-associated *Methanocorpusculum* phylogroups, with 6 host-associated clades represented by HQ MAGs and isolate genomes. With wombat and mahogany gliders representing phylogenetically and geographically unique organisms (Black et al., 2012; May-Collado et al., 2015), it may be expected that the *Methanocorpusculum* recovered from these species would be similarly phylogenetically distinct. However, genomes recovered from Chinese rhesus

macaque (*Macaca mulatta*) cluster closely with the wombat-associated *Methanocorpusculum* genomes (H. Wang et al., 2019), along with two MQ MAGs recovered from elephants (Ilmberger et al., 2014). This may suggest that the wombat-associated clade (HQ clade 6) represents a phylogroup of *Methanocorpusculum* with a wide host distribution. Host diversity was not limited to terrestrial animals, with a phylogenetically unique MAG recovered from the metagenome of a sperm whale (Li et al., 2019). Indeed, a subsequent study has also shown *Methanocorpusculum*, along with *Methanosphaera*, *Methanosarcina*, and *Methanomassiliicoccales* in wild baleen and toothed whales (Glaeser et al., 2021). Additionally, species of ptarmigan also provided a unique cluster of *Methanocorpusculum* as one of only two non-mammalian host (Salgado-Flores et al., 2019). Interestingly though, this chicken genomes clustered closer with the marsupial genome than the ptarmigan, despite both being derived from avian hosts. Together I have provided a substantial expansion of available *Methanocorpusculum* genomes from various host animals and geographical locations, allowing for a greater potential to characterise these novel genera of methanogenic archaea.

My bioinformatics analyses show that the HA *Methanocorpusculum* MAGs/genomes possess novel genes. Interestingly, the wombat-associated genomes were enriched for predicted phosphoenolpyruvate phosphomutase, phosphonopyruvate decarboxylase, and an alcohol dehydrogenase gene with predicted phosphonoacetaldehyde reductase activity. These three genes have been implicated in the biosynthesis of phosphonate compounds, with the gene cluster of *Streptomyces luridus* characterised in the production of the antibiotic dehydrophos, suggesting this lineage of *Methanocorpusculum* may also produce antibiotic phosphonate compounds (Shao, Blodgett et al. 2008). These antimicrobial compounds may provide the *Methanocorpsculum* with increased fitness and persistence in the GIT and allow for modulation of the surrounding microbes. Between the marsupial clades, the mahogany glider genomes were enriched for a wider variety of genes associated with nitrogen assimilation. *Methanocorpusculum* sp. MG contained all subunits of the nitrogenase molybdenum-iron protein (*nifDHK*), as well as *nifN* and nitrogen regulatory protein PII 1 and 2. Comparatively, *Methanocorpusculum* sp. CW153 contained only the *nifH* subunits. *nifH* can form homodimeric dinitrogenase reductase where *nifD* and *nifK* form dinitrogenase reductase as a hetero-tetramer, suggesting *Methanocorpusculum* sp. MG may encode for more efficient nitrogen assimilation through two different isotypes of dinitrogenase reductase (Burén et al., 2019; Einsle et al., 2002). *Methanocorpusculum* sp. CW153 was enriched for 7,8-dihydropterin-6-yl-methyl-4-(beta-D-ribofuranosyl)aminobenzene 5'-phosphate synthase, which produces the precursor for 5,6,7,8-tetrahydromethanopterin (H_4MPT) biosynthesis (Xu et al., 1999). Additionally, HA clades were also enriched for 5,6,7,8-tetrahydromethanopterin hydro-lyase (*fae*), a necessary component for methylene-H_4MPT formation and formaldehyde detoxification (Vorholt, Marx et al. 2000, Marx,

Chistoserdova et al. 2003). Interestingly, the marsupial genomes were significantly enriched for a gene identified as Benzoyl-CoA reductase subunit A (*badF*). This gene catalyses intermediate steps in benzoate degradation and may suggest the marsupial genomes can utilise aromatic compounds such as benzoate, phenol, or vanillin (Porter & Young, 2014). Interestingly, the predicted marsupial BadF showed high homology to the 2-hydroxyglutaryl-CoA dehydratase of *G. purinilytica* involved in amino acid fermentation, suggesting aromatic amino acids such as phenylalanine may be viable substrates (Porter & Young, 2014; Schweiger et al., 1987).

Analysis of the cazymes also showed the enrichment of GT, CE and AA enzymes in the marsupial-associated *Methanocorpusculum* clades. Despite the close phylogenetic association of the marsupial clades, my analyses suggest both contain a unique potential for glycosylation and likely produce unique cell surface structures adaptive for their host environment. AA6 genes were found in all *Methanocorpusculum* genomes, with a greater copy number observed in the marsupial-associated lineages. Typically identified as 1,4-Benzoquinone reductase, this gene is produced by *Moraxella* sp. for the metabolism of 4-nitrophenol, along with other quinoid compounds, such as 2-dimethoxybenzoquinone and vanillic acid, and can also induce the production of the quinone reductase (Akileswaran et al., 1999; Spain & Gibson, 1991). Similarly, 2-hydroxy-1,4-benzoquinone reductase is essential to the metabolism of 4-aminophenol by *Burkholderia* sp. strain AK-5 (Takenaka et al., 2011). This provides further potential evidence for phenol degradation, in addition to the identified *badF*. Interestingly, the third AA6 gene of *Methanocorpusculum* sp. MG showed a high similarity to a flavodoxin family protein of *Clostridium saccharobutylicum* (70.05%), potentially representing a recent acquisition by this clade of *Methanocorpusculum* for aminobenzoate degradation. Collectively this may suggest the host-associated environment contain benzene, phenol and quinoid compounds that the *Methanocorpusculum* have adapted to detoxify. However, future functional analyses are necessary to confirm this and to determine what metabolic burder they have on the different species.

Glycosyltransferases contained the largest variety of cazymes annotations, with GT66 and GT83 found in most *Methanocorpusculum* genomes. GT66 activities include dolichyl-diphosphooligosaccharide—protein glycotransferase (EC 2.4.99.18) and undecaprenyl-diphosphooligosaccharide—protein glycotransferase (EC 2.5.99.19)(Das & Heath, 1980; Maita et al., 2010). Archaeal oligosaccharyltransferase, encoded by archaeal glycosylation B (*aglB*), has previously displayed substantial diversity, providing a potentially wide range of N- and O-glycosylation to adapt to their diverse environments (Abu-Qarn & Eichler, 2007; Eichler, 2003). Similarly, GT83 glycosylates lipid A in *E. coli* and *Salmonella* with 4-amino-4-deoxy-l-arabinose (l-Ara4N)(Trent et al., 2001). Thus, these genes likely also play an important role in the glycosylation

of the *Methanocorpusculum* cell membrane. Interestingly, the glycosylation of lipid A with l-Ara4N provides resistance to polymyxin in *E. coli* and *Salmonella*, though it is unclear if a similar effect could be observed for *Methanocorpusculum* (Trent et al., 2001). Both marsupial clades contained GT8, however, the mahogany glider genomes contained an increased copy number. Genes with GT8 annotations have been characterised as lipopolysaccharide galactosyl/glucosyl/glucuronyltransferases and associated with O-antigen glycosylation (Lombard, Golaconda Ramulu et al. 2014). Additionally, alpha-1,3-rhamnosyltransferase, perosamine synthetase, and dTDP-4-dehydrorhamnose 3,5-epimerase were all found in the mahogany genomes and likely involved in the production and modification of outer membrane oligosaccharide biosynthesis (Christendat et al., 2000; Nakano et al., 2000; Zhao et al., 2007). These differences in glycosyltransferase genes likely represent host adaptations, with the genes enriched in the host-associated *Methanocorpusculum* likely required for persistent colonisation and adherence to the GIT.

In conclusion, I have expanded our understanding of gut methanogen diversity, and confirmed that the family *Methanocorpusculaceae* is a novel but widely prevalent lineage that colonises the gastrointestinal tract of herbivores, and particularly monogastric animals. I have achieved this via the isolation of two novel isolates from native Australian herbivores and the recovery of more than 134 additional MAGs. Using those resources, I have undertaken an in-depth comparison of the genetic potential inherent to the lineages of host- and environmental- origins. I propose the name *Methanocorpusculum vombatium*, sp. nov. (Type strain CW153^T) and *Methanocorpusculum petaurusium*, sp. nov. (Type strain MG^T) for the novel *Methanocorpusculum,* isolated from the Southern hairy nosed wombat and mahogany glider, respectively. While the analyses I presented in this Chapter primarily focused on the *Methanocorpusculaceae* lineages, I also successfully cultured a novel isolate of the family *Methanomethylophilaceae* from a mahogany glider*,* which is only the second cultured representative of the genus 'UBA_71'; as well as a representative of *M. gottschalkii* from the faecal sample of an Eastern grey kangaroo. These isolates and MAGs will support a deep functional understanding of the gut archaeome in both domesticated and wild herbivores, from which I hope to further identify host-specific adaptation and genetic targets unique to these novel methanogens, and to allow for the targeted manipulation the methanogen community. I briefly touch on these opportunities and future perspectives in the final Chapter of my thesis.

4.5 Published research article on work carried out in Chapter 4

A publication is being prepared on the isolation and characterisation of *Methanocorpusculum vombatium* and *Methanocorpusculum petaurusium*, along with the comparative genomics of the recovered *Methanocorpusculum* MAGs.

The isolation and characterisation of *Methanomethylophilus* sp. MG2 and *Methanobrevibacter gottschalkii* EGK will be prepared and published in separate publications.

Chapter 5: General Discussion

Methanogenic archaea occupy a unique biological niche within the GIT of diverse animal species and maintain 'microbial homeostasis' through the removal of end products generated from enteric fermentation (Nakamura et al., 2010; Sieber et al., 2012). In recent years, there has been a growing focus on archaea and the role different species play in GIT health and disease, as well as GIT ecology in non-human animals. Enabled by the increased efficiency and availability of DNA/RNA sequencing technologies, different animal species display diverse and unique 'archaeomes', which provide unique metabolic implications for each microbial community. Here, I have used bioinformatics approaches to identify and characterise novel phylogroups of methanogenic archaea from human and other vertebrate hosts. I also used culture-based approaches to isolate novel methanogen lineages from diverse animal hosts to bring these novel genomes to life. As a result of these efforts, I believe by Ph.D. research provides significant novel insights into the diversity of host-associated methanogenic archaea.

Chapter 2 described my successful isolation of *M. smithii* and *M. stadtmanae* from healthy Australian subjects and their capacity to tolerate and metabolise bile salts. My isolation of *M. stadtmanae* PA5 is only the second derived from humans and helped enable a comparative genomic analysis of the *Methanosphaera* genus (Hoedt et al., 2018). *M. smithii* and *M. stadtmanae* have both been reported to possess key genes in bile salt metabolism (*bsh* and 7/12α/b-*hsdh*) (Doden et al., 2018; Heinken et al., 2019; Kisiela et al., 2012) after Jones et al. (2008) had earlier shown the recombinant *bsh* of *M. smithii* to deconjugated tauro- and glyco-deoxycholic acid, when expressed in *E. coli*. My research in this Chapter is the first to show how bile acids affected the *in vitro* growth kinetics of multiple *M. smithii* and *M. stadtmanae* strains, with differential effects between and within lineages, and particularly for the *M. stadtmanae* strains. Interestingly though, *M. stadtmanae* isolate PA5 was shown to be unaffected by bile acids, where the type strain DSMZ3091^T showed a significant decrease in growth rate with increasing concentration. Considering these observations, I showed that while *bsh* expression is constitutive across all the strains, there were key differences in *bsh* transcript abundance in response to increasing bile salt concentrations. Additionally, the post-growth metabolome analysis of the respective cultures confirmed the deconjugation of glyco- and tauro-conjugated bile acids shown by Jones et al. (2008) and suggests that bile salt metabolism by *M. stadtmanae* DSMZ3091^T is distinct, with greater amounts of hyodeoxycholic acid produced compared to the other strains. As such, *M. stadtmanae* DSMZ3091^T may be able to dehydroxylate hyocholic acid to hyodeoxycholic acid, as previously described in isolate HDCA-1 recovered from the intestine of a rat (Eyssen et al., 1999). In fact, hyodeoxycholic acid was implicated in the suppression of intestinal epithelial cell proliferation through the FXR-dependent inhibition of the porcine IPEC-J2 cells (Song et al., 2020).

Interestingly, this study also showed dietary supplementation of hydeoxycholic acid in weaned piglets significantly increased the relative abundance of *Parabacteroides, prevotellaceae, campylobacter* and the *Clostridiales* genus *FamilyXIIITCG-001* compared to controls. A separate study of rats also showed a significantly altered bile acid content in a rat non-alcohol fatty liver disease (NAFLD) model, with a correlation shown between decreased hydeoxycholic acid and *Bacteroidetes* (Tang et al., 2019). These studies show hyodeoxycholic acids are able to modulate host cells but also correlate with a variety of microbial species. As such, future studies should address how the bile acid metabolite profiles produced by these archaea might translate into inter-Domain alterations in the gut microbiota. One way to examine this might be through the supplementation of bacterial cultures with methanogen culture supernatants grow with bile acids. This would allow for the exposure of bacteria of interest to the bile acid metabolites produced by the methanogens. Alternatively, co-culture studies could be conducted with methanogen and bacteria species, to show their interaction in the presence of bile acids. Although my analysis of polar metabolites recovered 13 bile acids, many bile acid derivatives were not detected, such as chenodeoxycholic acid and the associated conjugates. This bile acid is a major component of human bile acids accounting for ~40%, equal to that of cholic acid (~40%) and more than deoxycholic acid at ~20% (Chiang, 2017). Interestingly, chenodeoxycholic acid and its conjugates are present in Oxoid bile acid preparations, albeit at lower concentrations compared to cholic and deoxycholic acid (Hu et al., 2018). As such, further investigating the interactions between methanogens and important bile acids within the human bile acid pool through optimised detection and culture studies is essential to accurately capture the bile acid metabolite profiles that are produced and understand how these could in turn modulate the surrounding microbial community and host cells.

As part of my studies presented in Chapter 2, I revealed and confirmed that *M. stadtmanae* PA5 *bsh* contains a single nucleotide insertion, which is retained within *bsh* mRNA and thereby requires a +1 frameshift during translation to produce a full length Bsh product. My analysis of the *bsh* locus in the MAGs and isolate genomes provided evidence for this phenomenon across the *Methanobrevibacter* and *Methanosphaera* lineages. Translational read-through via novel tRNA for pyrrolysine (Antonov et al., 2013; Borrel et al., 2014) and -1 frameshifting in *Methanocaldococcus* and *Methanococcus spp.* for the synthesis of magnesium chelatase has previously been proposed (Antonov et al., 2013).Interestingly, programmed ribosomal frameshifting seems to be a common occurrence in many different viruses (Penn et al., 2020). +1 frameshifting has been proposed for the myovirus of halophilic archaea *Halorubrum sodomense* and *Haloarcula sinaiiensis* (Pietilä et al., 2013; Senčilo et al., 2013). The expression of Gag-Pol in HIV-1 has been shown to be regulated by programmed -1 ribosomal frameshifting, although a recent study also showed the host derived factor 'Shiftless' to inhibit the ribosomal frame shifting of HIV-1, as well as other viruses and cellular genes (X. Wang

et al., 2019). In terms of bacteria, frameshifting has been implicated in the regulation of several virulence associated genes. In *M. tuberculosis*, an insertion mutation in the β subunit of RNA polymerase is supressed by +1 translational frameshift resulting in a three amino acid alteration and provides increase rifampicin resistance (Huseby et al., 2020). Similarly, the iron uptake genes of *N. meningitidis* contains a stretch of guanines, not dissimilar to *M. stadtmanae* PA5, resulting in frameshifting that circumvents downstream nonsense codons (Richardson & Stojiljkovic, 1999). These programmed frameshifts have been speculated to provide proteome variation and increase the potential for adaptation (Fan et al., 2017). As such, the *bsh* frameshift of *M. stadtmanae* PA5 may represent a beneficial adaptation to the gut environment. This frameshifting may allow for greater regulation of gene expression, producing a greater percentage of the non-functional transcript under low bile salt concentrations and the complete transcript in the presence of bile salts (Advani & Dinman, 2016; Ketteler, 2012). It is worth noting that the primers used for the *bsh* expression analysis were specific to the first section of the truncated bile salt hydrolase gene and, as such, quantify the frequency at which the complete and truncated sequence are transcribed. The ribosomal slippage rate of different viruses is diverse, with 2 to 5% efficiency for the GagPol gene of retroviruses like HIV and 74 to 82% efficiency for genes of Theiler's murine encephalomyelitis virus (Atkins et al., 2016; Finch et al., 2015). Therefore, further transcriptional studies on the efficiency of the PA5 *bsh* transcriptional frame shift are necessary to determine if the increase in *bsh* translates to an increase in complete Bsh proteins.

Chapter 3 focused on my efforts to use bioinformatics tools to holistically examine the diversity and prevalence of human methanogenic archaea in health and disease, with a specific focus on the genus *Methanobrevibacter*. Coincident with my own efforts that recovered of 74 methanogen MAGs, Pasolli et al. (2019) produced more than 500 methanogen MAGs from human metagenome datasets produced from stool samples. Using these MAGs, as well as those draft and finished genomes produced from axenic isolates of human methanogenic archaea (Borrel, O'Toole, et al., 2013; Fricke et al., 2006; Hansen et al., 2011; Samuel et al., 2007), I created a database containing 512 HQ, human-derived genomes. While *M. smithii* remained the overwhelmingly predominant lineage represented in the database, my analyses provided key evidence that shows the *M. smithii* lineage is comprised of two groups with distinct phylogeny, hereafter referred to as *M. smithii* and *M. smithii*_A. The phylogenetic separation between these lineages extended to unique gene profiles for *M. smithii* and *M. smithii*_A, and in particular genes putatively involved with metal uptake and transport, with iron complex transport genes (*ABC.FEV.P, ABC.FEV.S, ABC.FEV.A*) significantly enriched in *M. smithii*_A genomes, and genes for a putative molybdate transport system (*modABC*) enriched in the *M. smithii* genomes. Interestingly, all MAGs recovered from CD patients in my analysis were

classified as *M. smithii*_A and MAGs recovered from T2D patients were all classified as *M. smithii*. Comparatively, MAGs recovered from patients with UC were classified as both *M. smithii* and *M. smithii*_A. Iron and molybdenum uptake is a key feature of many archaea, the former required for ubiquitous redox-related processes and the latter involved with C1 metabolism via formate dehydrogenases and environmental nitrogen fixation (Boyd et al., 2011; Johnson et al., 2021; Zhang & Gladyshev, 2008). Additionally, both iron acquisition and molybdenum-containing enzymes have also been implicated in bacterial virulence, including CD-associated adherent-invasive *E. coli* strain LF82 and extraintestinal pathogenic *E. coli* isolated from patients with IBD (Miquel et al., 2010; Petersen et al., 2009; Zhong et al., 2020). Additionally, poor management of T2D has been associated with the reduction of essential trace elements through increased urinary excretion (Khan & Awan, 2014). As such, the differential enrichment of metal uptake associated genes may represent specific adaptations for the metal starved ecology associated with different diseases and could differentially enrich of the *M. smithii* phylogroups.

My analyses also showed the *M. smithii* subgroup was enriched for a homolog of *acuB*. Although not well described in archaea, inactivation of *acuB* significantly reduced the growth of *Bacillus subtilis* using acetoin and butanediol (Grundy, Waters et al. 1993). According to Samuel, Hansen et al. (2007), acetate assimilation has been suggested as the primary carbon source for *M. smithii*; however, my analysis suggests members of the *M. smithii* phylogroup can de differentiated by their possession of genes involved with acetoin or butanediol utilisation and may explain the increased comparative abundance of *M. smithii*. My analyses also showed significant variation in genes associated with membrane glycosylation and cell surface structures for *M. smithii* and *M. smithii*_A. Together, the differential genotypes of *M. smithii* and *M. smithii*_A, along with the phylogenetic separation, warrants the identification of *M. smithii*_A as a distinct phylogroup containing a unique subset of genes. Thus, the differentiation of *M. smithii* into two distinct phylogroups is essential in understanding shifts observed in *M. smithii* during different disease states.

At the beginning of the research for this Chapter, I had anticipated that the genus *Methanosphaera,* which has long been thought to be the second most prevalent methanogen in the human gut (Dridi, Henry, et al., 2012), would have been more prevalent among the datasets. This data was primarily based on detection by PCR and culture-based methods, prior to the identification of *Methanomassiliicoccales* as a human-associated methanogen. However, only seven genomes were assigned to *Methanosphaera* and on further analysis, I reassigned two of these genomes to novel phylogenetic lineages closely associated with *Methanosphaera* sp. RUG761 and *M. cuniculi*. Indeed, the composition of my database would suggest that *Methanomassiliicoccales*-associated lineages are more prevalent and/or abundant and that, in addition to *Methanomassiliicoccus* and

Methanomethylophilus, the uncharacterised genus 'UBA71' is quite prevalent. This increased prevalence suggests the relative abundance of *Methanosphaera* is smaller than that of the major *Methanomassiliicoccales*-associated lineages in stool samples. Further MGS with increased sequencing depth should be conducted to provide additional support for the differential abundance of these methylotrophic methanogen lineages. Additionally, focus on the mucosal-associated methanogens would help to determine if this greater abundance of *Methanomassiliicoccales* remains current across the luminal and mucosal-associated communities, as well as different gastrointestinal sampling sites. Interestingly, no MAGs taxonomically classified as *M. luminyensis* were recovered, despite this isolate being the type strain and the first human isolate of *Methanomassiliicoccus* (Dridi, Fardeau, et al., 2012). Based on these results, I propose that genera 'UBA71' and *Methanomethylophilus* may in fact provide a more accurate representation of the dominant *Methanomassiliicoccales* within the human gut microbiome. Currently there are no cultured representatives of genus 'UBA71', therefore future work should focus on the cultured isolation of this genus, along with further axenic isolation of *Methanomethylophilus*, to allow for an in-depth functional and comparative analysis of these dominant human-associated *Methanomassiliicoccales* lineages.

Chapter 4 represents a substantial shift in my research focus, triggered by the COVID-19 pandemic. My original aim was to characterise how gut methanogen ecology was affected in patients with functional constipation, as part of a placebo-controlled randomised trial investigating a probiotic as a corrective intervention for this disorder. As summarised by Dimidi et al. (2014), previous studies have shown the beneficial effect of probiotic intervention to alleviate gut transit time, and improve stool consistency and frequency. More recently, a randomized double-blind placebo-controlled trial showed rifaximin improved colonic transit time and reduced methane production in subjects with chronic constipation (Ghoshal et al., 2018; Shah et al., 2018). However, all clinical research was postponed until late 2020, precluding my involvement with this study during my Ph.D. program. Through consultation with my supervisory committee, I chose to redirect my attention to examining the microbial resources developed as part of an Australian Research Council Discovery Project led by Professor Phil Hugenholtz and involving members of my supervisory team (Professors Morrison and Tyson) via the Australian Centre of Ecogenomics. This project involved the metagenomic analysis of the gut microbiomes from native Australian herbivores (Shiffman et al., 2017), and I focused on the examination of the methanogen communities resident in these animals. This direction allowed my research on methanogenic archaea to continue and increase the diversity of animal species from which I was able to isolation and characterise methanogens. In Chapter 2 and 3, I identified genetic adaptations in different groups of methanogens from humans. Chapter 4 allowed

me to widen the scope of these analyses to identify host-specific adaptations in non-human animals, further expanding our understanding of genetic differences between the different species of these under-characterised organisms. This provided a better understanding of why certain methanogen species dominate the archaeal community and what implications this has for the wider microbial community and animal host.

Previous analyses of the microbiome of Australian marsupials have primarily focused on the use of 16S rRNA sequencing with bacteria-specific or universal primers. This use of bacteria-specific primers causes inherent biases against archaeal sequences, restricting the detection of any archaeal sequences (Barker et al., 2013; Burke et al., 2018; Cheng et al., 2015; Eisenhofer et al., 2021). Although studies that utilised universal 16S rRNA primers on the microbiome of koala and kangaroo also failed to identify archaeal species suggesting any potential archaea population in these marsupials to be low in abundance (Brice et al., 2019; Dahlhausen et al., 2018; Gulino et al., 2013). In contrast, my analysis using universal amplicon-based sequencing detected *Methanobrevibacter* and two *Methanocorpusculum* spp., with differential enrichment of across the 23 marsupial species. Intriguingly, my analysis showed *Methanocorpusculum* as a predominant representative of the marsupial methanogen community. This was mirrored by my MGS analyses, which again showed *Methanocorpusculum* as a dominant member of the wombat samples. Typically, *Methanobrevibacter* and *Methanosphaera* have been studied as the dominant methanogens of the macropodid microbiome (Evans, 2011; Hoedt et al., 2016; Klieve et al., 2012). However, my analyses suggest *Methanocorpusculum* represents a dominant methanogen lineage in other marsupial species, namely wombat and mahogany gliders, while only being detected at a low abundance in the Macropodid microbiome.

In order to characterise the unique metabolic implications of this genus, I produced two independent axenic isolates assigned to the genus *Methanocorpusculum* from the stool samples of the common wombat (*Vombatus ursinus*) and mahogany glider (*Petaurus gracilis*), representing the first to be cultured from animal hosts. These isolates (and corresponding MAGs from the herbivore datasets) represent a new lineage within the *Methanocorpusculaceae* family. I also successfully isolated a representative of *M. gottschalkii* from an Eastern grey kangaroo (*Macropus giganteus*) and a novel *Methanomethylophilaceae* from a mahogany glider. Interestingly, the *Methanomethylophilaceae* isolate MG2 is a cultured representative of the currently uncharacterised 'UBA71' genus of *Methanomassiliicoccales*. These isolates substantially expand the number of available cultured methanogens from native Australian herbivores, including the recently isolated *Methanosphaera* sp. WGK6 from the forestomach content of a Western grey kangaroo (Hoedt et al., 2016) and *Methanobrevibacter* sp. WBY1 from a tammar wallaby (Evans, 2011).

Targeted isolation of novel methanogens is essential to improve both functional and bioinformatic characterisation of methanogen communities. My analysis of the marsupial metagenomic sequencing data utilised the GTDB-tk database containing 1,248 unique archaeal reference clusters and MAGs generated from the marsupial metagenomic sequencing as reference (Chaumeil et al., 2019). However, additional methanogens were detected from the metagenomic sequencing data when the genomes of the four cultured marsupial isolates were also used as reference genomes. As can be seen in Figure 5.1, *Methanocorpusculum* sp. MG was detected in samples from mahogany glider, squirrel glider, striped possum, and Lumholtz's tree-kangaroo, previously shown to contain no detectable methanogens using the standard reference database. Similarly, *Methanomethylophilus* sp. MG2 was also recovered from mahogany glider and squirrel glider, as well as Southern hairy-nosed wombats.

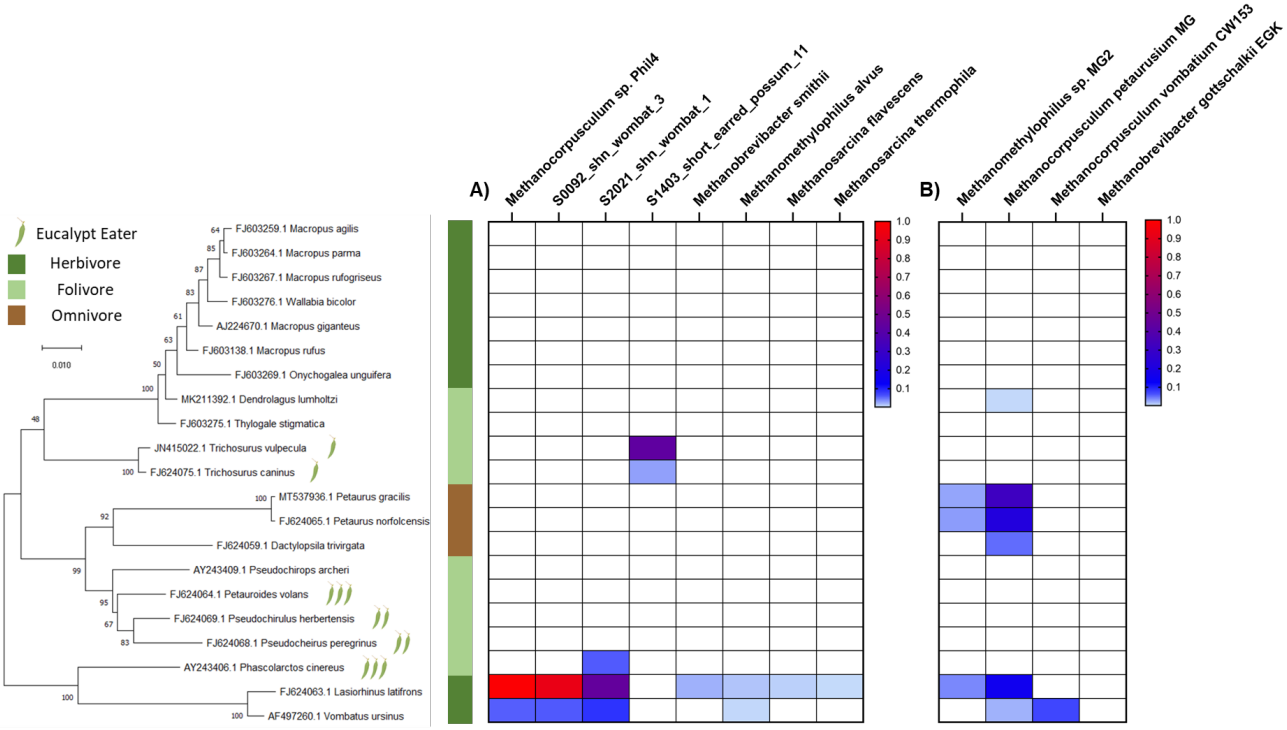


Figure 5.1. Methanogen profiles detected in marsupial species by metagenomic sequencing with different reference genomes. The phylogenetic tree was built using the von Willebrand factor (*vWF*) of respective species available from the NCBI nucleotide database. MEGA-X with MUSCLE was used to align the genes, with maximum-likelihood and 1000 bootstraps used for phylogeny. CoverM (v0.6.0) was used to detected methanogens based on the sequence coverage of reference genomes. **A)** used reference genomes from the GTDB-tk database and MAGs recovered from the marsupial samples. **B)** used the novel marsupial isolate genomes as reference. The host diet composition is displayed as per legend, reworked from Shiffman et al. (2017). The novel marsupial isolate genomes improved the detection of *Methanocorpusculum* in mahogany glider, squirrel glider, striped possum and Lumholtz's tree-kangaroo, as well as *Methanomethylophilus* in mahogany glider and squirrel glider.

Interestingly, *Methanocorpusculum* sp. CW153 was only detected in the common wombat samples, despite *Methanocorpusculum* being detected in both wombat species. As such, the *Methanocorpusculum* species may be differentially enriched in the common and Southern hairy-nosed wombat. *M. gottschalkii* was already represented in the GTDB-tk database as *M. gottschalkii* strains PG and HO, and was not detected in any marsupial sample. Curiously, *M. gottschalkii* EGK was also not detected in any marsupial sample, including the individual Eastern grey kangaroo sample from which it was recovered. This suggests the marsupial faecal samples require increased sequencing depth for accurate detection of the methanogen communities, particularly with regards to the kangaroo samples. However, the use of my cultured isolates resulted in the detection of methanogens in four marsupial species previously devoid of methanogens, including the detection of *Methanocorpusculum* sp. MG in Lumholtz's tree-kangaroo. As such, my results show the importance of the continued isolation of novel methanogen species to help augment current bioinformatic analyses.

Additionally, the isolation of novel methanogens is essential to ascribe functional outcomes to bioinformatic findings. Previous characterisation of *Methanosphaera* sp. WGK6 showed the propensity to use ethanol as a primary substrate for methanol-dependent methylotrophic methanogenesis (Hoedt et al., 2016). Previous analyses of *M. parvum* and *M. bavaricum* also showed a similar capacity to use 2-propanol and 2-butanol, in addition to hydrogenotrophic methanogenesis with carbon dioxide and hydrogen (Bleicher et al., 1989; Zellner et al., 1989). Indeed, characterisation of the *Methanocorpusculum adhs* showed the additional potential to utilise cyclopentanol, 2,3-butanediol, ethanol, and 1-propanol (Bleicher & Winter, 1991). However, my analysis showed that homologs of the previously identified *M. parvum* propanol-utilising *adh* were not present in any HQ host-associated *Methanocorpusculum* genome, including both *Methanocorpusculum* sp. CW153 and MG, suggesting they may be unable to utilise alcohols. Indeed, I was unable to show that neither *Methanocorpusculum* sp. CW153 nor MG utilised short-chain alcohols as a primary substrate for methanogenesis in the presence of CO_2 and H_2. Interestingly, this suggests that the host-associated *Methanocorpusculum* show a comparatively limited substrate utilisation profile compared to environmental strains, as well as the ethanol-utilising *Methanosphaera* sp. WGK6. This may suggest the environment-associated species have a greater availability of short-chain alcohols and thus have adapted the ability to utilise them for methanogenesis. It is worth noting that short-read sequencing was used to produce the genomes of the *Methanocorpusculum* (and other isolates through my analyses). This limits the confidence to accurate determine if genes are absent from the complete genomes or absent due to the presence of contig breaks and sequencing artifacts. Despite this, the large number of genomes used in my comparative genomic analyses provides statistical power to

support my findings. Additionally, the functional characterisation in the case of the *Methanocorpusculum* also suggests the *adh* gene is indeed absent from the cultured isolates. In the future, long-read sequencing, or a combination of long and short, should be used to produce an increased quality of genome assemblies.

Chapter 4 also focused on the recovery of *Methanocorpusculum* MAGs to increase the genomic representatives and allow for larger comparative analyses. Along these lines, I successfully recovered 134 *Methanocorpusculum* MAGs from publicly available metagenomes from 13 host species. Including my isolate genomes and MAGs generated from the marsupial samples, my analyses provided 142 new *Methanocorpusculum* genomes, accounting for 90% of all available genomes of the *Methanocorpusculum* genus. Collectively, phylogenetic analysis of the *Methanocorpusculum* showed at least 18 host-associated phylogroups, with 15 of these novel clades produced through my analyses. Concurrent my analyses, Nayfach et al. (2019) produced *Methanocorpusculum* MAGs from human faecal samples, which clustered with ruminant *Methanocorpusculum* MAGs. Interestingly, genomes recovered from Chinese rhesus macaque (*Macaca mulatta*) closely clustered with the wombat-associated *Methanocorpusculum* genomes (H. Wang et al., 2019), along with two MQ MAGs recovered from elephants (Ilmberger et al., 2014). This suggests the wombat-associated clade represents a phylogroup of *Methanocorpusculum* with a wide host distribution, despite marsupials representing a phylogenetically and geographically unique group of animals (Black et al., 2012; May-Collado et al., 2015).

Along with phylogenetic diversity, the *Methanocorpusculum* clades also showed a differential enrichment of genes. HA and Env genomes were differentially enriched for genes associated with methanogenesis, amino acid biosynthesis, and transport system proteins. One interesting observation was the enrichment of predicted phosphoenolpyruvate phosphomutase, phosphonopyruvate decarboxylase, and an alcohol dehydrogenase gene with predicted phosphonoacetaldehyde reductase activity in the wombat-associated genomes. Intriguingly, these genes were also differentially enriched between the *M. smithii* subgroups described in research Chapter 2. This cluster of genes has been characterised in the biosynthesis of phosphonate compounds such as the antibiotic dehydrophos produced by *Streptomyces luridus* and fosfomycin from *Streptomyces wedmorensis* (Nakashita et al., 1997; Shao et al., 2008; Woodyer et al., 2006). As such, this gene cluster may produce one or more phosphonate compounds that are differentially enriched across a wide variety of methanogen lineages, providing adaptation and host specificity by potentially modulating the surrounding microbial community. Additionally, the marsupial genomes were enriched for a gene identified as Benzoyl-CoA reductase subunit A, involved in benzoate degradation and may suggest the utilisation of aromatic compounds such as benzoate, phenol, or vanillin (Porter & Young, 2014). However, the

marsupial gene shows high homology to the 2-hydroxyglutaryl-CoA dehydratase of *Gottschalkia purinilytica* involved in amino acid fermentation, suggesting aromatic amino acids such as phenylalanine may be viable substrates (Porter & Young, 2014; Schweiger et al., 1987). To my understanding, there is currently no recognised methanogen that can utilise aromatic amino acids as a primary substrate for methanogenesis though utilisation has been shown through bacterial syntrophy (Chojnacka et al., 2015; Worm et al., 2010). Similar again to the *M. smithii* subgroups, cazymes were also differentially enriched between the different *Methanocorpusculum* clades. Specifically, glycosyltransferases were differentially enriched between the Env and HA clades, and within specific HA clades. These genes have been implicated in N- and O-glycosylation, as well as the production and modification of outer membrane oligosaccharide biosynthesis (Abu-Qarn & Eichler, 2007; Christendat et al., 2000; Eichler, 2003; Nakano et al., 2000; Zhao et al., 2007). This differential enrichment of glycosyltransferases is in part responsible for the production of cell membrane glycosylation for differential methanogen lineages and provides an adaptation to different host-specific environments.

In conclusion, methanogens have long been recognised as a key member of the gut microbial communities in animals and humans. While the advent of culture-independent techniques has greatly expanded our awareness of gut methanogenic archaea in recent years, these microbes have been relatively understudied in terms of their role(s) and involvement with host health and nutrition. My Ph.D. research has provided new insights into this important group of microbes via the combined use of bioinformatic and cultured-based (functional) studies, which includes the isolation of novel methanogen strains from human and non-human animal hosts. I believe my efforts and findings during my Ph.D. studies offer a significant contribution to improving our understanding of host-associated methanogen communities and pave the way for more detailed studies that better define genetic and functional traits relevant to host adaptation and specificity. From this foundation, new opportunities, and targets to productively manage populations of gut methanogens should be forthcoming, with potential benefits to medicine and agriculture.

Chapter 6: Appendix

Table 6.1. List of *bsh* used for the recovery of predicted *bsh* homologs from human methanogen genomes. *Bsh* sequences were downloaded from the NCBI Gene database. *Bsh* genes were used as a custom database with Kaptive (v0.5.1) to query available methanogen genomes.

Species	Genome	NCBI Refseq ID	NCBI Gene ID
M. smithii	*M. smithii* DSMZ861 (PS)	NC_009515.1	5216970
M. smithii	*M. smithii* KB11	NZ_CP017803.1	35119129
M. smithii	*M. smithii* C2	NZ_CAABOX000000000.1	-
M. smithii	*M. smithii* ACE6	NZ_CCXV00000000.1	-
M. smithii	*M. smithii* DSM 2374	NZ_ABYV00000000.2	-
M. smithii	*M. smithii* MGYG_HGUT_02163	NZ_CABMAD000000000.1	-
M. ruminantium	*M. ruminantium* M1	NC_013790.1	8771019
M. millerae	*M. millerae* SM9	NZ_CP011266.1	26736027
M. olleyae	*M. olleyae* YLM1	NZ_CP014265.1	28489153
M. stadtmanae	*M. stadtmanae* DSMZ3091 (MCB-3)	NC_007681.1	3854783
M. stadtmanae	*M. stadtmanae* MGYG_HGUT_02164	NZ_LR698975.1	-
M. stadtmanae	*M. stadtmanae* A6	NZ_MAIF00000000.1	-
***M.* sp. BMS**	*M.* sp. BMS	NZ_CP014213.1	-
M. formicicum	*M. formicium* BRM9	NZ_CP006933.1	24792731
M. alvus	*M. alvus* Mx1201	NC_020913.1	15140011
M. alvus	*M. alvus* Mx05	NZ_CP017686.1	38293231
M. alvus	*M. alvus* MGYG_HGUT_02456	NZ_LR699000.1	41321960

Table 6.2. List of polar metabolites detected in methanogen culture supernatants. Supernatant sample metabolites were analysed by LC-QToF, as per the methods.

STD1* 2-Methylcitric acid (C)	STD2* Uracil (C)	STD8-SM24* Elaidic acid
STD1* 4-Hydroxyproline	STD3* 2,3-Dihydroxybutanedioic acid	STD8-SM27* Allantoic acid
STD1* Adenine	STD3* 2-Ketobutyric acid (MH)	STD8-SM28* Stearic acid
STD1* cis-Aconitic acid	STD3* Adenosine monophosphate	STD8-SM30* D-Alanine
STD1* Citrulline	STD3* Alpha-Lactose	STD8-SM30* D-Lysine
STD1* D-2-Hydroxyglutaric acid	STD3* Betaine	STD8-SM32* 2-Methylbutyrylglycine
STD1* D-Glucurono-6,3-lactone	STD3* Cytidine monophosphate	STD8-SM7* N-Acetyl-L-alanine
STD1* D-Maltose	STD3* D-Fructose	STD9-SM11* Succinic anhydride
STD1* Fumaric acid	STD3* D-Ribose (MH)	STD9-SM16* Malonic acid
STD1* Gluconic acid	STD3* Glycyl-glycine	STD9-SM19* Threonic acid
STD1* Glucosamine	STD3* L-Alanine	STD9-SM25* D-Aspartic acid
STD1* Glycerophosphocholine	STD3* L-Norleucine	STD9-SM25* Oleic acid
STD1* Glycine	STD3* Phosphoric acid (B)	STD9-SM26* Pentadecanoic acid
STD1* Inosine	STD3* Pyruvic acid (MH)	STD9-SM27* Malic acid
STD1* L-Arginine	STD3* Ribitol	STD9-SM3* 3,4-Dihydroxyhydrocinnamic acid
STD1* L-Glutamic acid	STD3* Sorbitol	STD9-SM30* Hexadecanedioic acid
STD1* L-Leucine	STD4* Acetoacetic acid (MH)	STD9-SM30* Pelargonic acid
STD1* L-Lysine	STD4* Beta-Alanine (MH)	STD9-SM34* Adenosine 2',3'-cyclic phosphate
STD1* L-Methionine	STD4* Capric acid	STD9-SM35* Undecanedioic acid
STD1* L-Proline	STD4* Caproic acid	STD9-SM40* Glycocholic acid
STD1* L-Threonine	STD4* Caprylic acid	STD9-SM43* Sebacic acid
STD1* L-Valine	STD4* D-Glucose	STD9-SM6* L-Alloisoleucine
STD1* Mannitol	STD4* Galactonolactone (B)	STD10-SM12* Itaconic acid
STD1* Myristic acid	STD4* Galacturonic acid	STD10-SM15* Acetylglycine
STD1* Orotic acid	STD4* Heptadecanoic acid	STD10-SM17* N-Acetyl-L-tyrosine
STD1* Phthalic acid	STD4* Isocitric acid	STD10-SM21* Turanose
STD1* Sucrose 6-Phosphate	STD4* L-Phenylalanine	STD10-SM22* Alpha-Linolenic acid
STD1* Taurine	STD4* L-Tryptophan	STD10-SM27* N6-Acetyl-L-lysine
STD1* Uridine 5'-monophosphate	STD4* Methylmalonic acid	STD10-SM31* D-Glutamic acid
STD2* Alanylglycine	STD4* Pantothenic acid	STD10-SM37* scyllo-Inositol
STD2* Citric acid	STD4* Trehalose	STD10-SM8* Beta-Leucine
STD2* Cyclic AMP	STD4* Xanthine	STD11-SM14* Dimethylglycine
STD2* D-Erythrose 4-phosphate	STD5* Beta-Glycerophosphoric acid	STD11-SM19* D-Xylitol
STD2* D-Glucuronic acid	STD5* Cellobiose	STD11-SM20* Pimelic acid
STD2* Dodecanoic acid	STD5* D-Xylose (C)	STD11-SM20* Shikimic acid
STD2* Gamma-Aminobutyric acid	STD5* L-Gulonolactone	STD11-SM22* Dodecanedioic acid
STD2* Glycerol 3-phosphate	STD5* myo-Inositol	STD11-SM25* 2,6-Dihydroxybenzoic acid
STD2* Guanosine	STD6-SM1* Ethylmethylacetic acid	STD11-SM25* N-Acetyl-L-phenylalanine
STD2* L-Asparagine	STD6-SM11* 2-Hydroxyethanesulfonate	STD11-SM32* Ribothymidine
STD2* L-Aspartic acid	STD6-SM11* Pipecolic acid	STD11-SM34* L-Sorbose
STD2* L-Cystine	STD6-SM15* Galactitol	STD11-SM37* Hydrocinnamic acid
STD2* L-Homoserine	STD6-SM15* Galacturonic acid	STD11-SM4* 2-Hydroxy-2-methylbutyric acid
STD2* L-Isoleucine	STD6-SM2* 2-Furoylglycine	STD12-SM15* Lactulose

STD2* L-Malic acid	STD6-SM7* L-Allothreonine	STD12-SM19* Sarcosine
STD2* L-Serine	STD6-SM9* L-Arabitol	STD12-SM23* Suberic acid
STD2* L-Tyrosine	STD6-SM9* Methionine sulfoxide	STD12-SM28* Hypoxanthine
STD2* N-Acetylglutamine	STD7-SM10* 5-Methylcytidine	STD12-SM32* Indoxyl sulfate
STD2* N-Acetylornithine	STD7-SM16* Glyceric acid	STD12-SM33* Isovalerylglycine
STD2* N-Alpha-acetyllysine	STD7-SM18* Isomaltose	STD12-SM33* L-Cysteine
STD2* Ornithine	STD7-SM19* Pyrrolidonecarboxylic acid	STD12-SM33* Tetradecanedioic acid
STD2* Oxalic acid	STD7-SM25* Azelaic acid	STD12-SM36* 3-Methylphenylacetic acid
STD2* Palmitic acid	STD7-SM26* Vaccenic acid	STD12-SM42* 7a-Hydroxy-3-oxo-5b-cholanoic acid
STD2* Phenylacetylglycine	STD7-SM28* Cholic acid	STD12-SM5* Valeric Acid
STD2* Succinic acid	STD7-SM34* 3-Methyladipic acid	STD13-SM30* Octadecanedioic acid
STD2* Sucrose	STD7-SM35* myo-Inositol	STD13-SM35* Benzoic acid
STD2* Thiamine pyrophosphate	STD7-SM6* N-Acetylglutamic acid	STD13-SM8* L-Pipecolic acid
STD2* Thymine	STD8-SM23* Linoleic acid	

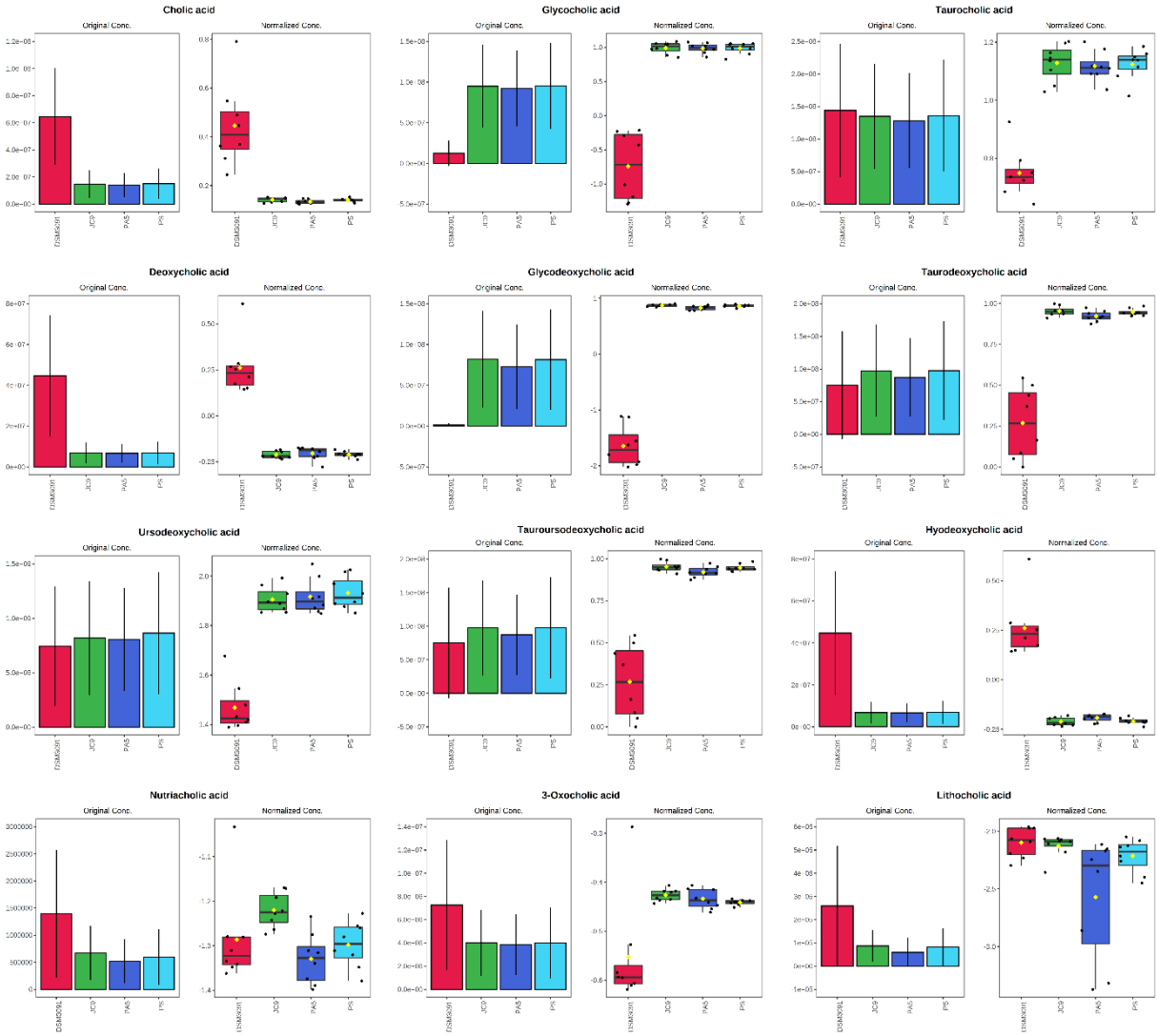


Figure 6.1. Graphs displaying the concentration of statistically significant bile acids between strains according to analysis by ANOVA. P values for respective bile acids: cholic acid - 3.2618E-8, glycocholic acid - 1.4802E-14, taurocholic acid - 6.4378E-13, deoxycholic acid - 2.9092E-13, glycodeoxycholic acid - 1.019E-22, taurodeoxycholic acid - 2.5378E-13, ursodeoxycholic acid - 1.9408E-13, tauroursodeoxycholic acid - 2.5491E-13, hyodeoxycholic acid - 2.2621E-13, nutriacholic - 0.026133, 3-oxocholic acid - 2.7574E-4, and lithocholic acid - 0.0096143. All bile acids were significantly different for *M. stadtmanae* DSMZ3091, except for nutriacholic and lithocholic which were significantly different for *M. smithii* JC9 and *M. stadtmanae* PA5, respectively. Uninoculated media metabolite concentrations were not removed from respective culture supernatants values.

Table 6.3. List of primers used for PCR and qRT-PCR. *Methanosphaera* (*Msp.*) BSH, *Methanobrevibacter* (*Mbb.*) BSH and *Methanosphaera* (*Msp.*) PA5 GAP primer pairs were designed as a part of this study, as per the methods.

Primer ID	Gene Target	Sequence (3'-5')	Annealing Temp. (°C)	Amplicon (bps)	Reference
***Msp.* BSH F1**	*M. stadtmanae bsh*	GGCTGGATTAAATTTCGCTGG	58	192	This study
***Msp.* BSH R1**		TGAAGTGGTGAAAGGGGAAGT			This study
***Mbb.* BSH F5**	*M. smithii bsh*	CCAGTTGCCTCCATCTCCAC	56	83	This study
***Mbb.* BSH R5**		CCAGTCCTTCCTCTACAACCTC			This study
***Msp.* PA5 GAP F1**	*M. stadtmanae* PA5 *bsh* truncation	TGGCTGGATTAAATTTCGCTGG	58	671	This study
***Msp.* PA5 GAP R1**		ACAGCATTTATCTGAGAGTTTCCA			This study
958A (F)	Archaeal 16S rRNA	AATTGGAKTCAACGCCGG	60	142	DeLong, 1992
1100Ar (R)		TGGGTCTCGCTCGTTG			Whitford et al., 2001
86F	Archaeal 16S rRNA	GCTCAGTAACACGTGG	58	1254	Wright and Pimm, 2003
1340R		CGGTGTGTGCAAGGAG			
27F	Bacterial 16S rRNA	AGAGTTTGATCCTHHCTCAG	58	1465	Enticknap et al., 2006
1492R		GGTTACCTTGTTACGACTT			

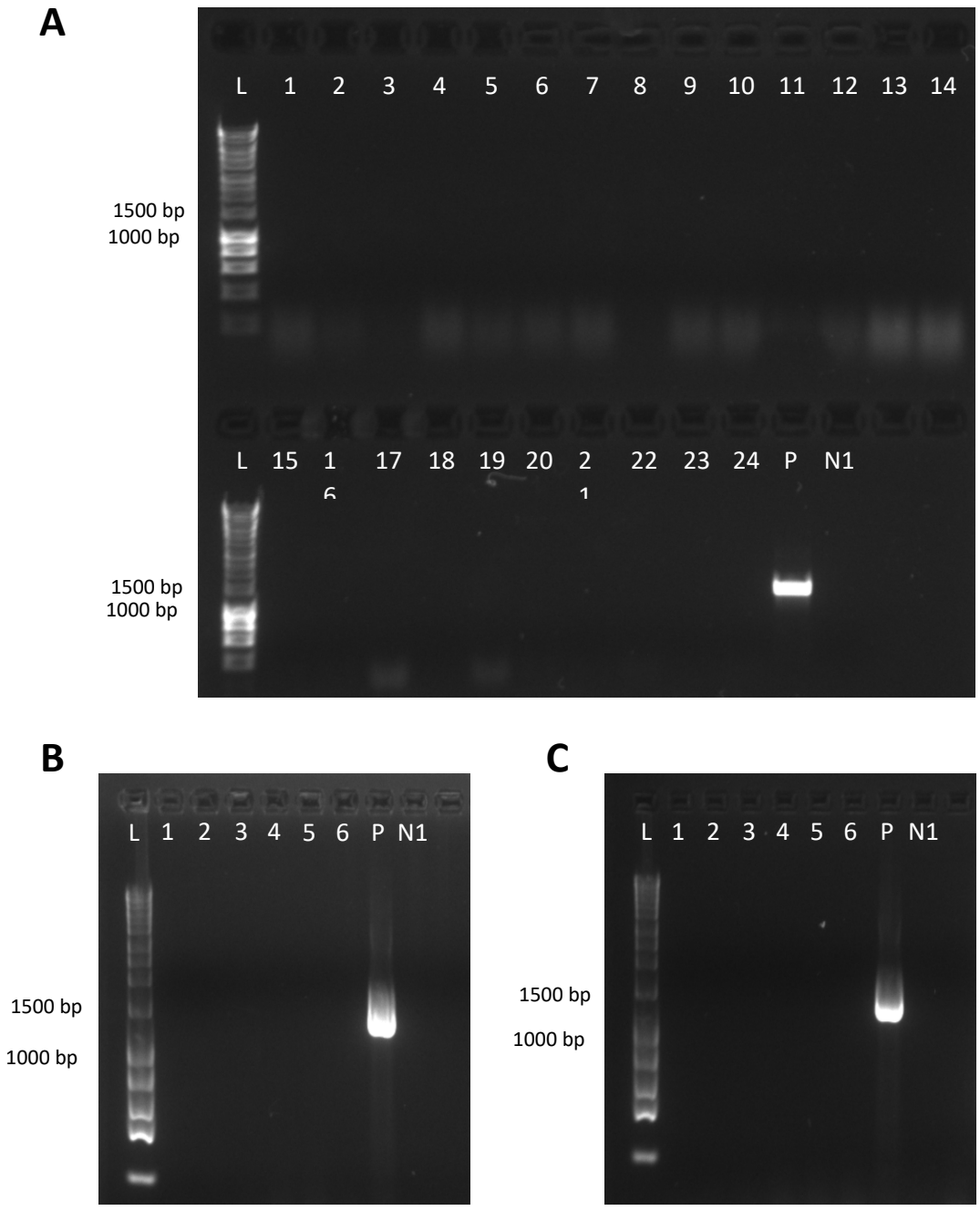


Figure 6.2. Gel electrophoresis showing no amplification of gDNA in RNA samples. RNA samples were used as template in archaeal (86F/1340R) PCR to test for gDNA contamination. **A)** Lanes 1-12 shows *M. stadtmanae* DSMZ3091 0.0, 0.1, 1.0% RNA extractions in biological and technical duplicate. Lanes 13-24 shows *M. stadtmanae* PA5 0.0, 0.1, 1.0% RNA extractions in biological and technical duplicate. **B)** and **C)** show *M. smithii* PS and *M. smithii* JC9 respectively, with lanes 1-6 representing 0.0, 0.1, 1.0% RNA extractions in technical duplicate. Lane P shows amplification with gDNA of the respective strain as the positive control. Lane N1 shows amplification of UdH_2O as a negative control. Lane L shows 1 Kbp Bioline HyperLadder. No gDNA was detected in any of the RNA samples.

Table 6.4. Concentration and quality values for *M. stadtmanae* DSMZ3091 and PA5 RNA extractions. Samples were run on three Bioanalyser RNA Nano Chips (Agilent Technologies), as per the manufacturer's instructions. Samples are denoted as bile acid concentration (0, 0.1 or 1.0%), biological replicated (one or two) and technical replicate (one or two). Three samples were rerun due to initial failed runs. All samples produced high quality (RIN=>9) RNA that was used for further analyses.

Chip/sample No.	Sample	RNA Conc. (ng/µL)	rRNA Ratio (23s/16s)	RIN
Chip 1 Sample 1	PA5 0-1-1	356	1.6	9.1
Chip 1 Sample 2	PA5 0-1-2	186	1.5	9.5
Chip 1 Sample 3	PA5 0-2-1	50	1.3	9.5
Chip 1 Sample 4	PA5 0-2-2	228	1.5	9.5
Chip 1 Sample 5	PA5 0.1-1-1	239	1.3	9.2
Chip 1 Sample 6	PA5 0.1-1-2	299	1.4	9.2
Chip 1 Sample 7	PA5 0.1-2-1	300	1.3	9.3
Chip 1 Sample 8	PA5 0.1-2-2	403	1.4	9.5
Chip 1 Sample 9	PA5 1.0-1-1	402	1.5	9.6
Chip 1 Sample 10	PA5 1.0-1-2	348	1.4	9.6
Chip 3 Sample 1	PA5 1.0-2-1	273	1.7	9.8
Chip 1 Sample 12	PA5 1.0-2-2	381	1.6	9.7
Chip 2 Sample 1	3091 0-1-1	569	1.7	9.8
Chip 2 Sample 2	3091 0-1-2	581	1.7	9.8
Chip 2 Sample 3	3091 0-2-1	319	1.7	9.9
Chip 2 Sample 4	3091 0-2-2	294	1.7	9.9
Chip 2 Sample 5	3091 0.1-1-1	1273	1.7	9.8
Chip 3 Sample 2	3091 0.1-1-2	107	2.1	9.6
Chip 3 Sample 3	3091 0.1-2-1	783	1.8	9.9
Chip 2 Sample 8	3091 0.1-2-2	307	2	9.7
Chip 2 Sample 9	3091 1.0-1-1	153	1.6	9.8
Chip 2 Sample 10	3091 1.0-1-2	302	1.5	9.6
Chip 2 Sample 11	3091 1.0-2-1	88	1.6	9.7
Chip 2 Sample 12	3091 1.0-2-2	229	1.8	9.8

Figure 6.3. Multiple alignment of *bsh* nucleotide sequences from representative genomes *bsh* clusters. Nucleotide sequences were aligned using MUSCLE in MEGA-X (Kumar et al., 2018).

Table 6.5. List of reference genomes included in comparative analyses. Genomes were downloaded the from NCBI genome database. *Methanobrevibacter smithii* JC9, *Methanobrevibacter smithii* PAM and *Methanocorpusculum* sp. CW153 are cultured isolates recovered by our group.

Genome ID	Genome	Genome Assembly	Source	GTDB Taxonomic Classification
Mbb_87_7	*Methanobrevibacter* sp. 87.7	GCA_002208625.1	Sheep	s__Methanobrevibacter_B sp002208625
Mbb_A27	*Methanobrevibacter* sp. A27	GCA_001729385.1	-	s__Methanobrevibacter_A gottschalkii
Mbb_A54	*Methanobrevibacter* sp. A54	GCA_001729455.1	-	s__Methanobrevibacter_A smithii
Mbb_AbM4	*Methanobrevibacter* sp. AbM4	GCA_000404165.1	Sheep	s__Methanobrevibacter_B boviskoreani
Mbb_arboriphilus_ANOR1	*Methanobrevibacter arboriphilus* ANOR1	GCA_000513315.1	Human	s__Methanobrevibacter_C arboriphilus_A
Mbb_arboriphilus_DH1	*Methanobrevibacter arboriphilus* DH1	GCA_002072215.1	Environmental	s__Methanobrevibacter_C arboriphilus
Mbb_arboriphilus_JCM9315	*Methanobrevibacter arboriphilus* JCM9315	GCA_001315885.1	Environmental	s__Methanobrevibacter_C arboriphilus_A
Mbb_bovisboreani_JH1	*Methanobrevibacter boviskoreani* JH1	GCA_000320505.2	Bovine	s__Methanobrevibacter_B boviskoreani
Mbb_curvatus_DSM11111	*Methanobrevibacter curvatus* DSM11111	GCA_001639295.1	Termite	s__Methanobrevibacter_D curvatus
Mbb_cuticularis_DSM11139	*Methanobrevibacter cuticularis* DSM11139	GCA_001639285.1	Termite	s__Methanobrevibacter_C cuticularis
Mbb_filiformis_DSM11501	*Methanobrevibacter filiformis* DSM11501	GCA_001639265.1	Termite	s__Methanobrevibacter_D filiformis
Mbb_gottschalkii	*Methanobrevibacter gottschalkii* PG	GCA_900109595.1	Pig	s__Methanobrevibacter_A gottschalkii
Mbb_millerae_SM9	*Methanobrevibacter millerae* SM9	GCA_001477655.1	Bovine	s__Methanobrevibacter_A millerae
Mbb_NOE	*Methanobrevibacter* sp. NOE	GCA_003315655.1	Environmental	s__Methanobrevibacter_C sp003315655
Mbb_olleyae_YLM1	*Methanobrevibacter olleyae* YLM1	GCA_001563245.1	Sheep	s__Methanobrevibacter olleyae
Mbb_oralis_DSM7256	*Methanobrevibacter oralis* DSM7256	GCA_001639275.1	Human	s__Methanobrevibacter_A oralis
Mbb_oralis_JMR01	*Methanobrevibacter oralis* JMR01	GCA_000529525.1	Human	s__Methanobrevibacter_A oralis
Mbb_ruminantium_M1	*Methanobrevibacter ruminantium* M1	GCA_000024185.1	Bovine	s__Methanobrevibacter ruminantium
Mbb_smithii_ACE6	*Methanobrevibacter smithii* ACE6	GCA_000824705.1	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_C2	*Methanobrevibacter smithii* C2 CSUR P5816	GCA_900650605.1	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_CAG186	*Methanobrevibacter smithii* CAG186	GCA_000437055.1	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_DSM2374	*Methanobrevibacter smithii* DSMZ2374	GCA_000151225.1	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_DSM2375	*Methanobrevibacter smithii* DSMZ2375	GCA_000151245.1	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_KB11	*Methanobrevibacter smithii* KB11	GCA_002813085.1	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_PS	*Methanobrevibacter smithii* PS	GCA_000016525.1	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_strain_JC9	*Methanobrevibacter smithii* JC9	-	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_strain_PAM	*Methanobrevibacter smithii* PAM	-	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_TS145A	*Methanobrevibacter smithii* TS145A	GCA_000189795.2	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_TS145B	*Methanobrevibacter smithii* TS145B	GCA_000189815.2	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_TS146A	*Methanobrevibacter smithii* TS146A	GCA_000189835.2	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_TS146B	*Methanobrevibacter smithii* TS146B	GCA_000189855.2	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_TS146C	*Methanobrevibacter smithii* TS146C	GCA_000189875.2	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_TS146D	*Methanobrevibacter smithii* TS146D	GCA_000189895.2	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_TS146E	*Methanobrevibacter smithii* TS146E	GCA_000189915.2	Human	s__Methanobrevibacter_A smithii_A
Mbb_smithii_TS147A	*Methanobrevibacter smithii* TS147A	GCA_000189935.2	Human	s__Methanobrevibacter_A smithii_A
Mbb_smithii_TS147B	*Methanobrevibacter smithii* TS147B	GCA_000189955.2	Human	s__Methanobrevibacter_A smithii_A
Mbb_smithii_TS147C	*Methanobrevibacter smithii* TS147C	GCA_000189975.2	Human	s__Methanobrevibacter_A smithii_A
Mbb_smithii_TS94A	*Methanobrevibacter smithii* TS94A	GCA_000189995.2	Human	s__Methanobrevibacter_A smithii_A
Mbb_smithii_TS94B	*Methanobrevibacter smithii* TS94B	GCA_000190015.2	Human	s__Methanobrevibacter_A smithii_A
Mbb_smithii_TS94C	*Methanobrevibacter smithii* TS94C	GCA_000190035.2	Human	s__Methanobrevibacter_A smithii_A
Mbb_smithii_TS95A	*Methanobrevibacter smithii* TS95A	GCA_000190055.2	Human	s__Methanobrevibacter_A smithii_A

Mbb_smithii_TS95B	*Methanobrevibacter smithii* TS95B	GCA_000190075.2	Human	s__Methanobrevibacter_A smithii_A
Mbb_smithii_TS95C	*Methanobrevibacter smithii* TS95C	GCA_000190095.2	Human	s__Methanobrevibacter_A smithii_A
Mbb_smithii_TS95D	*Methanobrevibacter smithii* TS95D	GCA_000190115.2	Human	s__Methanobrevibacter_A smithii_A
Mbb_smithii_TS96A	*Methanobrevibacter smithii* TS96A	GCA_000190135.2	Human	s__Methanobrevibacter_A smithii_A
Mbb_smithii_TS96B	*Methanobrevibacter smithii* TS96B	GCA_000190155.2	Human	s__Methanobrevibacter_A smithii_A
Mbb_smithii_TS96C	*Methanobrevibacter smithii* TS96C	GCA_000190175.2	Human	s__Methanobrevibacter_A smithii
Mbb_smithii_WWM1085	*Methanobrevibacter smithii* WWM1085	GCA_002252585.1	Human	s__Methanobrevibacter_A smithii_A
Mbb_thaueri_DSM11995	*Methanobrevibacter thaueri* DSM11995	GCA_003111625.1	Bovine	s__Methanobrevibacter_A thaueri
Mbm_veterum_MK4	*Methanobacterium veterum* MK4	GCA_000745485.1	Environmental	s__Methanobacterium_D veterum
Mbb_woesei_DSM11979	*Methanobrevibacter woesei* DSM11979	GCA_003111605.1	Goose	s__Methanobrevibacter_A woesei
Mbb_wolinii_SH	*Methanobrevibacter wolinii* SH	GCA_000621965.1	Sheep	s__Methanobrevibacter_B wolinii
Mbb_YE315	*Methanobrevibacter* sp YE315	GCA_001548675.1	Bovine	s__Methanobrevibacter_A sp001548675
Mbm_A39	*Methanobacterium bryantii* A39	GCA_001729285.1	Environmental	s__Methanobacterium_D bryantii
Mbm_arcticum_M2	*Methanobacterium veterum* M2	GCA_000746075.1	Environmental	s__Methanobacterium_D veterum
Mbm_BAmetb5	*Methanobacterium* sp. BAmetb5	GCA_003491305.1	Environmental	s__Methanobacterium sp003491305
Mbm_BRmetb2	*Methanobacterium* sp. BRmetb2	GCA_003491285.1	Environmental	s__UBA117 sp002494785
Mbm_bryantii_MoH	*Methanobacterium bryantii* MoH	GCA_002287175.1	Environmental	s__Methanobacterium_D bryantii
Mbm_congolense	*Methanobacterium congolense* Buetzberg	GCA_900095295.1	Environmental	s__Methanobacterium_C congolense
Mbm_formicicum_BRM9	*Methanobacterium formicicum* BRM9	GCA_000762265.1	Bovine	s__Methanobacterium formicicum
Mbm_formicicum_DSM3637	*Methanobacterium formicicum* DSM3637	GCA_000302455.1	Environmental	s__Methanobacterium formicicum_A
Mbm_lacus_AL21	*Methanobacterium lacus* AL21	GCA_000191585.1	Environmental	s__Methanobacterium_B lacus
Mbm_MB1	*Methanobacterium* sp. MB1	GCA_000499765.1	Bovine	s__Methanobacterium sp000499765
Mbm_MZA1	*Methanobacterium* sp. MZA1	GCA_002813675.1	Environmental	s__Methanobacterium subterraneum
Mbm_paludis_SWAN1	*Methanobacterium paludis* SWAN1	GCA_000214725.1	Environmental	s__Methanobacterium_C paludis
Mbm_SMA27	*Methanobacterium* sp. SMA27	GCA_000744455.1	Environmental	s__Methanobacterium_B sp000744455
Mbm_subterraneum_MOMB1	*Methanobacterium subterraneum* MO-MB1	GCA_002813655.1	Environmental	s__Methanobacterium subterraneum
Mcl_bourgensis_MS2	*Methanoculleus bourgensis* MS2	GCA_000304355.2	Environmental	s__Methanoculleus bourgensis
Mcl_CAG1088	*Methanoculleus* sp. CAG1088	GCA_000437835.1	Human	s__Methanomethylophilus alvus
Mcl_chikugoensis_L21	*Methanoculleus chikugoensis* L21-II-0	GCA_900095385.1	Environmental	s__Methanoculleus chikugoensis_A
Mcl_EBM46	*Methanoculleus* sp. EBM46	GCA_001896715.1	Environmental	s__Methanoculleus sp001896715
Mcl_horonobensis_T10	*Methanoculleus horonobensis* T10	GCA_001602375.1	Environmental	s__Methanoculleus horonobensis
Mcl_MAB1	*Methanoculleus* sp. MAB1	GCA_900036045.1	Environmental	s__Methanoculleus bourgensis
Mcl_marisnigri_JR1	*Methanoculleus* sp. JR1	GCA_000015825.1	Environmental	s__Methanoculleus marisnigri
Mcl_MH98A	*Methanoculleus* sp. MH98A	GCA_000691865.1	Environmental	s__Methanoculleus sp000691865
Mcl_sediminis_S3Fa	*Methanoculleus sediminis* S3Fa	GCA_001017125.1	Environmental	s__Methanoculleus sediminis
Mcl_taiwanesis_CYW4	*Methanoculleus taiwanensis* CYW4	GCA_004102725.1	Environmental	s__Methanoculleus_A taiwanensis
Mcl_thermophilus_CR1	*Methanoculleus thermophilus* CR1	GCA_001571405.1	Environmental	s__Methanoculleus thermophilus
Mcp_CW153	*Methanocorpusculum* sp. CW153	-	Wombat	s__Methanocorpusculum sp001940805
Methanogenic_archaeon_ISO4-G1	Methanogenic archaeon ISO4-G1	GCA_001563305.1	Sheep	s__ISO4-G1 sp001563305
Mmp_alvus_Mx05	*Methanomethylophilus* alvus Mx05	GCA_003711245.1	Human	s__Methanomethylophilus alvus
Mmp_avlus_Mx1201	*Methanomethylophilus* alvus Mx1201	GCA_000300255.2	Human	s__Methanomethylophilus alvus
Mms_BRNA1	Thermoplasmatales archaeon BRNA1	GCA_000350305.1	Bovine	s__Methanomethylophilus sp000350305
Mms_intestinalis_Mx1	*Methanomassiliicoccus intestinalis* Issoire-Mx1	GCA_000404225.1	Human	s__Methanomassiliicoccus intestinalis
Mms_luminyensis_B10	*Methanomassiliicoccus luminyensis* B10	GCA_000308215.1	Human	s__Methanomassiliicoccus luminyensis
Msp_BMS	*Methanosphaera* sp. BMS	GCA_003268005.1	Bovine	s__Methanosphaera sp003268005
Msp_cuniculi_IR-7	*Methanosphaera cuniculi* IR-7	GCA_002287195.1	Rabbit	s__Methanosphaera cuniculi

Msp_DEW79	*Methanosphaera stadtmanae* DEW79	-	Human	s__Methanosphaera stadtmanae
Msp_PA5	*Methanosphaera stadtmanae* PA5	GCA_003265405.1	Human	s__Methanosphaera stadtmanae
Msp_rholeuAM130	*Methanosphaera* sp. rholeuAM130	GCA_003266065.1	Bovine	s__Methanosphaera sp003266065
Msp_rholeuAM270	*Methanosphaera* sp. rholeuAM270	GCA_003266165.1	Bovine	s__Methanosphaera sp003266165
Msp_rholeuAM6	*Methanosphaera* sp. rholeuAM6	GCA_003266105.1	Bovine	s__Methanosphaera sp003266105
Msp_rholeuAM74	*Methanosphaera* sp. rholeuAM74	GCA_003266075.1	Bovine	s__Methanosphaera sp003266075
Msp_SHI1033	*Methanosphaera* sp. SHI1033	GCA_003266175.1	Sheep	s__Methanosphaera sp002509095
Msp_SHI613	*Methanosphaera* sp. SHI613	GCA_003266145.1	Sheep	s__Methanosphaera sp003266145
Msp_stadtmanae_DSM3091	*Methanosphaera stadtmanae* DSM3091	GCA_000012545.1	Human	s__Methanosphaera stadtmanae
Msp_WGK6	*Methanosphaera* sp. WGK6	GCA_001729965.1	Kangaroo	s__Methanosphaera sp001729965

Table 6.6. Table displaying KEGG Orthologs (KO) enriched between *M. smithii* and *M. smithii*_A genomes recovered from T2D and CD samples. EnrichM (v0.4.9) was used to determined differentially enriched KOs between CD genomes (n=3) and T2D genomes (n=8). KO annotations with corrected P values ≤0.05 were included. No KO reached significance with p value correction.

KO annotation	T2D Genomes	CD Genomes	P value	Corrected p value	Gene description
K02020	8	0	0.006060606	0.74469697	modA; molybdate transport system substrate-binding protein
K02017	8	0	0.006060606	0.74469697	modC; molybdate transport system ATP-binding protein [EC:3.6.3.29]
K11741	0	3	0.006060606	0.74469697	sugE; quaternary ammonium compound-resistance protein SugE
K03651	8	0	0.006060606	0.74469697	cpdA; 3',5'-cyclic-AMP phosphodiesterase [EC:3.1.4.53]
K10947	8	0	0.006060606	0.74469697	padR; PadR family transcriptional regulator, regulatory protein PadR
K03276	0	3	0.006060606	0.74469697	waaR, waaT, rfaJ; UDP-glucose/galactose:(glucosyl)LPS alpha-1,2-glucosyl/galactosyltransferase [EC:2.4.1.-]
K02018	8	0	0.006060606	0.74469697	modB; molybdate transport system permease protein
K03626	8	0	0.006060606	0.74469697	EGD2, NACA; nascent polypeptide-associated complex subunit alpha
K00721	7	0	0.024242424	1	DPM1; dolichol-phosphate mannosyltransferase [EC:2.4.1.83]
K03100	7	0	0.024242424	1	lepB; signal peptidase I [EC:3.4.21.89]
K04767	7	0	0.024242424	1	acuB; acetoin utilization protein AcuB
K07494	1	3	0.024242424	1	K07494; putative transposase

Table 6.7. Table displaying KEGG Orthologs (KO) enriched between *M. smithii* and *M. smithii*_A genomes recovered from UC and CD samples. EnrichM (v0.4.9) was used to determined differentially enriched KOs between CD genomes (n=3) and UC genomes (n=22). KO annotations with corrected P values ≤0.05 were included. No KO reached significance with p value correction.

KO annotation	CD Genomes	UC Genomes	P value	Corrected p value	Gene Description
K11741	3	2	0.005646527	1	sugE; quaternary ammonium compound-resistance protein SugE
K03651	0	18	0.005646527	1	cpdA; 3',5'-cyclic-AMP phosphodiesterase [EC:3.1.4.53]
K10947	0	18	0.005646527	1	padR; PadR family transcriptional regulator, regulatory protein PadR
K03626	0	18	0.005646527	1	EGD2, NACA; nascent polypeptide-associated complex subunit alpha
K07494	3	2	0.005646527	1	K07494; putative transposase
K03276	3	3	0.011293055	1	waaR, waaT, rfaJ; UDP-glucose/galactose:(glucosyl)LPS alpha-1,2-glucosyl/galactosyltransferase [EC:2.4.1.-]
K06943	1	20	0.011857708	1	NOG1; nucleolar GTP-binding protein
K07038	1	20	0.011857708	1	K07038; inner membrane protein
K07979	2	0	0.011857708	1	ytrA; GntR family transcriptional regulator
K13993	1	20	0.011857708	1	HSP20; HSP20 family protein
K06921	3	4	0.019762846	1	K06921; uncharacterized protein
K04767	0	16	0.019762846	1	acuB; acetoin utilization protein AcuB
K03100	0	15	0.031620553	1	lepB; signal peptidase I [EC:3.4.21.89]
K00721	0	14	0.04743083	1	DPM1; dolichol-phosphate mannosyltransferase [EC:2.4.1.83]
K02020	0	14	0.04743083	1	modA; molybdate transport system substrate-binding protein
K02017	0	14	0.04743083	1	modC; molybdate transport system ATP-binding protein [EC:3.6.3.29]
K02018	0	14	0.04743083	1	modB; molybdate transport system permease protein

Table 6.8. KEGG Ortholog (KO) annotations enriched in *M. smithii* and *M. smithii*_A phylogroups. EnrichM (v0.4.9) was used to determined KOs enriched in *M. smithii*_A (n=84) and *M. smithii* (n=365). KO annotations with corrected P values ≤0.05 were included. Corrected significance values showed 12 KO annotations enriched in *M. smithii* and 20 enriched in *M. smithii*_A.

KO	Smithii	Smithii_A	Pvalue	Corrected Pvalue	Gene Description
K11741	0	82	1.45E-88	3.04E-85	sugE; quaternary ammonium compound-resistance protein SugE
K10947	358	0	1.75E-83	1.83E-80	padR; PadR family transcriptional regulator, regulatory protein PadR
K01077	0	78	7.51E-81	5.25E-78	E3.1.3.1, phoA, phoB; v [EC:3.1.3.1]
K03651	352	0	1.01E-77	5.28E-75	cpdA; 3',5'-cyclic-AMP phosphodiesterase [EC:3.1.4.53]
K04767	336	0	1.64E-66	6.90E-64	acuB; acetoin utilization protein AcuB
K02018	337	1	1.07E-64	3.74E-62	modB; molybdate transport system permease protein
K02020	337	2	1.36E-62	4.07E-60	modA; molybdate transport system substrate-binding protein
K03100	332	1	7.13E-62	1.87E-59	lepB; signal peptidase I [EC:3.4.21.89]
K02017	331	1	2.45E-61	5.70E-59	modC; molybdate transport system ATP-binding protein [EC:3.6.3.29]
K00721	323	0	1.10E-59	2.31E-57	DPM1; dolichol-phosphate mannosyltransferase [EC:2.4.1.83]
K03276	21	74	9.97E-54	1.90E-51	waaR, waaT, rfaJ; UDP-glucose/galactose:(glucosyl)LPS alpha-1,2-glucosyl/galactosyltransferase [EC:2.4.1.-]
K02015	75	84	7.74E-47	1.35E-44	ABC.FEV.P; iron complex transport system permease protein
K02016	74	83	5.69E-45	9.17E-43	ABC.FEV.S; iron complex transport system substrate-binding protein
K03626	356	27	1.62E-41	2.43E-39	EGD2, NACA; nascent polypeptide-associated complex subunit alpha
K06921	97	83	3.09E-38	4.32E-36	K06921; uncharacterized protein
K04719	1	36	1.22E-28	1.60E-26	bluB; 5,6-dimethylbenzimidazole synthase [EC:1.13.11.79]
K02013	180	84	1.21E-22	1.49E-20	ABC.FEV.A; iron complex transport system ATP-binding protein [EC:3.6.3.34]
K08169	7	27	8.95E-16	1.04E-13	yebQ; MFS transporter, DHA2 family, multidrug resistance protein
K07499	35	38	6.03E-13	6.66E-11	K07499; putative transposase
K18909	1	15	5.50E-11	5.76E-09	mepR; MarR family transcriptional regulator, repressor for mepA
K10761	98	0	1.02E-10	1.01E-08	THG1; tRNA(His) guanylyltransferase [EC:2.7.7.79]
K19302	3	14	1.05E-08	9.97E-07	bcrC; undecaprenyl-diphosphatase [EC:3.6.1.27]
K07494	15	21	3.00E-08	2.74E-06	K07494; putative transposase
K07319	1	9	1.65E-06	0.00014443	yhdJ; adenine-specific DNA-methyltransferase [EC:2.1.1.72]
K09124	158	58	2.04E-05	0.001708448	K09124; uncharacterized protein
K03696	4	9	7.17E-05	0.005568769	clpC; ATP-dependent Clp protease ATP-binding subunit ClpC
K05847	361	75	7.17E-05	0.005568769	opuA; osmoprotectant transport system ATP-binding protein
K01156	46	26	0.00011199	0.00838728	res; type III restriction enzyme [EC:3.1.21.5]
K19425	299	81	0.000341789	0.024714854	epsH; glycosyltransferase EpsH [EC:2.4.-.-]
K19174	24	17	0.000447436	0.031275763	dptG; DNA phosphorothioation-dependent restriction protein DptG
K11085	5	8	0.000635347	0.042978176	msbA; ATP-binding cassette, subfamily B, bacterial MsbA [EC:3.6.3.-]
K01155	83	6	0.000735607	0.048205216	E3.1.21.4; type II restriction enzyme [EC:3.1.21.4]

Table 6.9. KEGG Ortholog (KO) annotations enriched in *M. smithii* subgroups. EnrichM (v0.4.9) was used to determined KOs enriched in 11 *M. smithii* subgroups. KO annotations with P values ≤0.05 were included.

KO	Gene ID	Smithii Subgroup 2 true	Smithii Subgroup 1 true	P value	Correct P value
K06221	dkgA	0	4	0.032955	1
K07499	K07499	0	4	0.032955	1
K07342	SEC61G, SSS1, secE	35	22	0.036005	1
K15342	cas1	27	14	0.046228	1
KO	Gene ID	Smithii Subgroup 3 true	Smithii Subgroup 1 true	P value	Corrected P value
K15342	cas1	19	14	0.000573	0.688180807
K10761	THG1	1	14	0.001347	0.808619591
K03658	helD	20	21	0.014782	1
K04096	smf	11	6	0.017368	1
K01155	E3.1.21.4	1	10	0.017467	1
K03427	hsdM	14	10	0.020862	1
K16264	czcD, zitB	20	22	0.032157	1
K00666	ACSF2	19	20	0.03365	1
K01657	trpE	15	28	0.034924	1
K01156	res	6	2	0.049676	1
K02864	RP-L10, MRPL10, rplJ	14	27	0.049676	1
K02867	RP-L11, MRPL11, rplK	14	27	0.049676	1
K02601	nusG	14	27	0.049676	1
KO	Gene ID	Smithii Subgroup 4 true	Smithii Subgroup 1 true	P value	Corrected P value
K15342	cas1	24	14	0.038168	1
KO	Gene ID	Smithii Subgroup 5 true	Smithii Subgroup 1 true	P value	Corrected P value
K02016	ABC.FEV.S	18	1	0.000734	0.910816419
K00002	AKR1A1, adh	32	28	0.001718	1
K02015	ABC.FEV.P	18	2	0.003	1
K00666	ACSF2	44	20	0.013767	1
K10761	THG1	10	14	0.021161	1
K19003	mgdA	5	9	0.033076	1
K00402	mcrG	41	29	0.040748	1
K03421	mcrC	41	29	0.040748	1
K00973	E2.7.7.24, rfbA, rffH	48	26	0.049952	1
K20444	rfbC	48	26	0.049952	1
K00003	hom	48	26	0.049952	1
K01710	E4.2.1.46, rfbB, rffG	48	26	0.049952	1
K01790	rfbC, rmlC	48	26	0.049952	1
KO	Gene ID	Smithii Subgroup 6 true	Smithii Subgroup 1 true	P value	Corrected P value
K02016	ABC.FEV.S	13	1	0.000431	0.545671839
K02015	ABC.FEV.P	13	2	0.002129	1

K01154	hsdS	24	12	0.00333	1
K15342	cas1	25	14	0.006107	1
K00721	DPM1	30	23	0.010543	1
K19091	cas6	1	8	0.012183	1
K19139	csm4	18	8	0.018244	1
K11646	K11646	30	24	0.023721	1
K16264	czcD, zitB	29	22	0.025691	1
K02013	ABC.FEV.A	21	12	0.036992	1
K03427	hsdM	19	10	0.037887	1
K07016	csm1, cas10	19	10	0.037887	1
KO	Gene ID	Smithii Subgroup 7 true	Smithii Subgroup 1 true	P value	Corrected P value
K15342	cas1	15	14	0.002896	1
K19175	dptH	5	0	0.003575	1
K19174	dptG	5	0	0.003575	1
K10761	THG1	1	14	0.006902	1
K03601	xseA	15	16	0.008212	1
K19115	csh2	4	0	0.012215	1
K19114	csh1	4	0	0.012215	1
K01338	lon	6	2	0.016682	1
K00666	ACSF2	16	20	0.016977	1
K19140	csm5	10	8	0.029799	1
K19139	csm4	10	8	0.029799	1
K07016	csm1, cas10	11	10	0.034518	1
K00963	UGP2, galU, galF	16	21	0.036594	1
K09131	K09131	16	21	0.036594	1
K02124	ATPVK, ntpK, atpK	16	21	0.036594	1
K16264	czcD, zitB	16	22	0.039763	1
K00077	panE, apbA	16	22	0.039763	1
K09721	K09721	16	22	0.039763	1
K02189	cbiG	16	22	0.039763	1
K02227	cbiB, cobD	16	22	0.039763	1
K19116	cas5h	4	1	0.046775	1
KO	Gene ID	Smithii Subgroup 1 true	Smithii Subgroup 8 true	P value	Corrected P value
K15342	cas1	14	32	0.00019	0.218077161
K00721	DPM1	23	35	0.006336	1
K09721	K09721	22	34	0.018629	1
K02189	cbiG	22	34	0.018629	1
K19139	csm4	8	20	0.023602	1
K07068	K07068	25	35	0.037381	1
K02927	RP-L40e, RPL40	25	35	0.037381	1
K20608	tet	25	35	0.037381	1
K19140	csm5	8	19	0.042887	1
K07016	csm1, cas10	10	22	0.043699	1
K09002	csm3	10	21	0.049174	1
KO	Gene ID	Smithii Subgroup 1 true	Smithii Subgroup 9 true	P value	Corrected P value

K02016	ABC.FEV.S	1	4	0.020212	1
K07075	K07075	0	3	0.020638	1
K00666	ACSF2	20	12	0.039183	1
KO	Gene ID	Smithii Subgroup 1 true	Smithii Subgroup 10 true	P value	Corrected P value
K02016	ABC.FEV.S	1	10	0.001641	1
K19115	csh2	0	7	0.003242	1
K02015	ABC.FEV.P	2	10	0.007617	1
K19114	csh1	0	6	0.007942	1
K15342	cas1	14	22	0.00964	1
K00666	ACSF2	20	25	0.012867	1
K00096	araM, egsA	29	21	0.018909	1
K00963	UGP2, galU, galF	21	25	0.026743	1
K09131	K09131	21	25	0.026743	1
K07487	K07487	2	8	0.034553	1
K19116	cas5h	1	6	0.043835	1
K13812	fae-hps	29	22	0.043835	1
K10761	THG1	14	5	0.045432	1
KO	Gene ID	Smithii Subgroup 1 true	Smithii Subgroup 11 true	P value	Corrected P value
K01841	pepM	0	18	3.97E-05	0.024305626
K09459	E4.1.1.82	0	18	3.97E-05	0.024305626
K19954	adh1	0	16	0.000108	0.0440977
K06921	K06921	11	5	0.009172	1
K07342	SEC61G, SSS1, secE	22	43	0.023836	1
K00666	ACSF2	20	41	0.025991	1
K19115	csh2	0	7	0.037962	1
K19114	csh1	0	7	0.037962	1
K02189	cbiG	22	42	0.041839	1
KO	Gene ID	Smithii Subgroup 3 true	Smithii Subgroup 2 true	P value	Corrected P value
K02867	RP-L11, MRPL11, rplK	14	37	0.001068	0.640868565
K02601	nusG	14	37	0.001068	0.640868565
K07342	SEC61G, SSS1, secE	12	35	0.002091	0.836396308
K03658	helD	20	25	0.004647	1
K02864	RP-L10, MRPL10, rplJ	14	36	0.005718	1
K02863	RP-L1, MRPL1, rplA	14	36	0.005718	1
K10761	THG1	1	14	0.010163	1
K07499	K07499	4	0	0.012266	1
K03100	lepB	15	36	0.016876	1
K06989	nadX, ASPDH	17	37	0.038961	1
K01872	AARS, alaS	17	37	0.038961	1
K11105	cvrA, nhaP2	17	37	0.038961	1
K01928	murE	17	37	0.038961	1
K04096	smf	11	10	0.047572	1
KO	Gene ID	Smithii Subgroup 4 true	Smithii Subgroup 2 true	P value	Corrected P value
K06921	K06921	6	18	0.011841	1

K12410	npdA	32	30	0.012662	1
K19427	epsJ	7	18	0.02574	1
KO	Gene ID	Smithii Subgroup 5 true	Smithii Subgroup 2 true	P value	Corrected P value
K02016	ABC.FEV.S	18	2	0.000578	0.344387525
K02015	ABC.FEV.P	18	2	0.000578	0.344387525
K19427	epsJ	7	18	0.000817	0.344387525
K00002	AKR1A1, adh	32	35	0.002513	0.794692808
K06188	aqpZ	10	0	0.00411	0.994320718
K06921	K06921	9	18	0.004716	0.994320718
K19425	epsH	45	26	0.006494	1
K07499	K07499	8	0	0.008643	1
K00402	mcrG	41	37	0.017003	1
K03421	mcrC	41	37	0.017003	1
K03701	uvrA	45	28	0.026551	1
K02428	rdgB	48	33	0.032618	1
K03496	parA, soj	48	33	0.032618	1
K03422	mcrD	42	37	0.033374	1
K00399	mcrA	42	37	0.033374	1
K03658	helD	42	25	0.033452	1
K07458	vsr	1	6	0.039611	1
K03655	recG	21	8	0.039698	1
KO	Gene ID	Smithii Subgroup 6 true	Smithii Subgroup 2 true	P value	Corrected P value
K02016	ABC.FEV.S	13	2	0.000274	0.174657108
K02015	ABC.FEV.P	13	2	0.000274	0.174657108
K01154	hsdS	24	16	0.00283	1
K02013	ABC.FEV.A	21	13	0.006753	1
K12410	npdA	30	30	0.01418	1
K06921	K06921	6	18	0.021151	1
K19003	mgdA	14	8	0.038481	1
KO	Gene ID	Smithii Subgroup 7 true	Smithii Subgroup 2 true	P value	Corrected P value
K19175	dptH	5	2	0.020869	1
K19174	dptG	5	2	0.020869	1
K10761	THG1	1	14	0.021993	1
K19115	csh2	4	1	0.024988	1
K19116	cas5h	4	1	0.024988	1
K19114	csh1	4	1	0.024988	1
K03100	lepB	12	36	0.024988	1
KO	Gene ID	Smithii Subgroup 2 true	Smithii Subgroup 8 true	P value	Corrected P value
K06921	K06921	18	8	0.028665	1
KO	Gene ID	Smithii Subgroup 2 true	Smithii Subgroup 9 true	P value	Corrected P value
K19427	epsJ	18	1	0.016713	1
K02013	ABC.FEV.A	13	9	0.02174	1
K02016	ABC.FEV.S	2	4	0.025737	1
K02015	ABC.FEV.P	2	4	0.025737	1
K02927	RP-L40e, RPL40	36	9	0.040755	1
KO	Gene ID	Smithii Subgroup 4 true	Smithii Subgroup 3 true	P value	Corrected P value

K03658	helD	21	20	0.003616	1
K07342	SEC61G, SSS1, secE	29	12	0.01396	1
K04096	smf	7	11	0.019276	1
K10761	THG1	10	1	0.03532	1
KO	Gene ID	Smithii Subgroup 5 true	Smithii Subgroup 3 true	P value	Corrected P value
K02864	RP-L10, MRPL10, rplJ	47	14	0.001999	0.623714273
K02867	RP-L11, MRPL11, rplK	47	14	0.001999	0.623714273
K02601	nusG	47	14	0.001999	0.623714273
K02863	RP-L1, MRPL1, rplA	47	14	0.001999	0.623714273
K00721	DPM1	35	20	0.0073	1
K00002	AKR1A1, adh	32	19	0.014513	1
K19425	epsH	45	14	0.015504	1
K01657	trpE	46	15	0.020039	1
K06989	nadX, ASPDH	48	17	0.022747	1
K11105	cvrA, nhaP2	48	17	0.022747	1
K00766	trpD	47	16	0.023797	1
K01609	trpC	47	16	0.023797	1
K01658	trpG	47	16	0.023797	1
K01695	trpA	47	16	0.023797	1
K06188	aqpZ	10	0	0.027624	1
K07342	SEC61G, SSS1, secE	41	12	0.028809	1
KO	Gene ID	Smithii Subgroup 6 true	Smithii Subgroup 3 true	P value	Corrected P value
K02864	RP-L10, MRPL10, rplJ	30	14	0.002439	0.764677453
K02867	RP-L11, MRPL11, rplK	30	14	0.002439	0.764677453
K02601	nusG	30	14	0.002439	0.764677453
K02863	RP-L1, MRPL1, rplA	30	14	0.002439	0.764677453
K19003	mgdA	14	2	0.011901	1
K19091	cas6	1	6	0.012418	1
K03658	helD	22	20	0.015468	1
K04096	smf	6	11	0.015483	1
K10761	THG1	11	1	0.016034	1
K01657	trpE	29	15	0.031709	1
K01154	hsdS	24	10	0.034182	1
K01156	res	2	6	0.046873	1
KO	Gene ID	Smithii Subgroup 7 true	Smithii Subgroup 3 true	P value	Corrected P value
K02864	RP-L10, MRPL10, rplJ	16	14	0.024011	1
K02867	RP-L11, MRPL11, rplK	16	14	0.024011	1
K02601	nusG	16	14	0.024011	1
K02863	RP-L1, MRPL1, rplA	16	14	0.024011	1
K07342	SEC61G, SSS1, secE	15	12	0.026049	1
K19003	mgdA	7	2	0.049137	1

KO	Gene ID	Smithii Subgroup 3 true	Smithii Subgroup 8 true	P value	Corrected P value
K02864	RP-L10, MRPL10, rplJ	14	35	0.001337	0.387403791
K02867	RP-L11, MRPL11, rplK	14	35	0.001337	0.387403791
K02601	nusG	14	35	0.001337	0.387403791
K02863	RP-L1, MRPL1, rplA	14	35	0.001337	0.387403791
K07342	SEC61G, SSS1, secE	12	33	0.00277	0.64206022
K04096	smf	11	6	0.005913	1
K03658	helD	20	26	0.019078	1
K06989	nadX, ASPDH	17	35	0.043453	1
K09384	K09384	3	0	0.043453	1
K07105	K07105	17	35	0.043453	1
KO	Gene ID	Smithii Subgroup 3 true	Smithii Subgroup 9 true	P value	Corrected P value
K03658	helD	20	8	0.013765	1
K07075	K07075	0	3	0.044355	1
KO	Gene ID	Smithii Subgroup 5 true	Smithii Subgroup 4 true	P value	Corrected P value
K00002	AKR1A1, adh	32	32	0.000104	0.125068789
K02435	gatC, GATC	47	24	0.002298	0.917702986
K07458	vsr	1	8	0.002298	0.917702986
K00721	DPM1	35	30	0.021345	1
K19425	epsH	45	24	0.022847	1
K03658	helD	42	21	0.026259	1
K00402	mcrG	41	32	0.03793	1
K19003	mgdA	5	10	0.037992	1
K01154	hsdS	29	11	0.03921	1
K02016	ABC.FEV.S	18	5	0.044477	1
K02015	ABC.FEV.P	18	5	0.044477	1
KO	Gene ID	Smithii Subgroup 6 true	Smithii Subgroup 4 true	P value	Corrected P value
K01154	hsdS	24	11	0.000361	0.438765398
K02016	ABC.FEV.S	13	5	0.024708	1
K02015	ABC.FEV.P	13	5	0.024708	1
K02013	ABC.FEV.A	21	14	0.044478	1
K03088	rpoE	4	0	0.049127	1
K09729	K09729	26	32	0.049127	1
KO	Gene ID	Smithii Subgroup 7 true	Smithii Subgroup 4 true	P value	Corrected P value
K19116	cas5h	4	1	0.036564	1
KO	Gene ID	Smithii Subgroup 4 true	Smithii Subgroup 8 true	P value	Corrected P value
K09384	K09384	6	0	0.00908	1
K02435	gatC, GATC	24	33	0.039306	1
K19175	dptH	4	0	0.046916	1
K07068	K07068	28	35	0.046916	1
KO	Gene ID	Smithii Subgroup 4 true	Smithii Subgroup 9 true	P value	Corrected P value
K06921	K06921	6	7	0.022594	1
KO	Gene ID	Smithii Subgroup 6 true	Smithii Subgroup 5 true	P value	Corrected P value
K00002	AKR1A1, adh	30	32	0.000262	0.342594923

K19003	mgdA	14	5	0.000764	0.500261421
K00721	DPM1	30	35	0.001174	0.512562217
K19091	cas6	1	13	0.007149	1
K19139	csm4	18	15	0.018307	1
K07075	K07075	4	0	0.019212	1
K19138	csm2	19	17	0.020537	1
K19090	cas5t	1	11	0.023752	1
K19075	cst2, cas7	1	11	0.023752	1
K19088	cst1, cas8a	1	11	0.023752	1
K01902	sucD	25	47	0.028943	1
K06188	aqpZ	1	10	0.043411	1
KO	Gene ID	Smithii Subgroup 5 true	Smithii Subgroup 7 true	P value	Corrected P value
K02435	gatC, GATC	47	11	0.002903	1
K19425	epsH	45	10	0.005521	1
K00002	AKR1A1, adh	32	16	0.006557	1
K19003	mgdA	5	7	0.00679	1
K19427	epsJ	7	8	0.006912	1
K19174	dptG	2	5	0.008569	1
K19114	csh1	1	4	0.012031	1
K02227	cbiB, cobD	34	16	0.013822	1
K19175	dptH	3	5	0.019237	1
K19115	csh2	2	4	0.030285	1
K19116	cas5h	2	4	0.030285	1
K19139	csm4	15	10	0.038691	1
KO	Gene ID	Smithii Subgroup 5 true	Smithii Subgroup 8 true	P value	Corrected P value
K00002	AKR1A1, adh	32	35	4.67E-05	0.056370795
K00721	DPM1	35	35	0.000443	0.267499825
K06188	aqpZ	10	0	0.004158	1
K02016	ABC.FEV.S	18	3	0.004201	1
K02015	ABC.FEV.P	18	3	0.004201	1
K19427	epsJ	7	15	0.005553	1
K07458	vsr	1	7	0.00878	1
K07075	K07075	0	5	0.011181	1
K19090	cas5t	11	1	0.011238	1
K09721	K09721	37	34	0.011238	1
K19075	cst2, cas7	11	1	0.011238	1
K02189	cbiG	37	34	0.011238	1
K19088	cst1, cas8a	11	1	0.011238	1
K07579	K07579	40	35	0.018347	1
K09384	K09384	8	0	0.018347	1
K19139	csm4	15	20	0.024761	1
K15342	cas1	34	32	0.027729	1
K06936	K06936	42	35	0.036812	1
K19138	csm2	17	21	0.043976	1
K19425	epsH	45	27	0.046306	1
K01154	hsdS	29	13	0.046534	1

K09002	csm3	18	21	0.048651	1
KO	Gene ID	Smithii Subgroup 5 true	Smithii Subgroup 9 true	P value	Corrected P value
K07075	K07075	0	3	0.006429	1
K06921	K06921	9	7	0.010241	1
K00002	AKR1A1, adh	32	12	0.025312	1
KO	Gene ID	Smithii Subgroup 6 true	Smithii Subgroup 7 true	P value	Corrected P value
K19115	csh2	0	4	0.011153	1
K19114	csh1	0	4	0.011153	1
K01338	lon	2	6	0.014715	1
K10761	THG1	11	1	0.035037	1
K19175	dptH	2	5	0.040201	1
K19174	dptG	2	5	0.040201	1
K19116	cas5h	1	4	0.043019	1
K19091	cas6	1	4	0.043019	1
K01154	hsdS	24	8	0.04824	1
K02016	ABC.FEV.S	13	2	0.048686	1
K02015	ABC.FEV.P	13	2	0.048686	1
KO	Gene ID	Smithii Subgroup 6 true	Smithii Subgroup 8 true	P value	Corrected P value
K01154	hsdS	24	13	0.000927	0.618390209
K02016	ABC.FEV.S	13	3	0.001508	0.618390209
K02015	ABC.FEV.P	13	3	0.001508	0.618390209
K02013	ABC.FEV.A	21	13	0.01254	1
K01902	sucD	25	35	0.017253	1
K07744	K07744	30	29	0.02684	1
K09153	K09153	30	29	0.02684	1
K19003	mgdA	14	7	0.033049	1
K06936	K06936	26	35	0.040478	1
K03088	rpoE	4	0	0.040478	1
K04771	degP, htrA	4	0	0.040478	1
K03442	mscS	26	35	0.040478	1
K09721	K09721	24	34	0.042434	1
K02189	cbiG	24	34	0.042434	1
KO	Gene ID	Smithii Subgroup 6 true	Smithii Subgroup 9 true	P value	Corrected P value
K19091	cas6	1	4	0.018388	1
K06921	K06921	6	7	0.026133	1
K19003	mgdA	14	1	0.030893	1
K19138	csm2	19	3	0.040041	1
K07016	csm1, cas10	19	3	0.040041	1
KO	Gene ID	Smithii Subgroup 7 true	Smithii Subgroup 8 true	P value	Corrected P value
K19175	dptH	5	0	0.001859	0.97993921
K19174	dptG	5	0	0.001859	0.97993921
K02435	gatC, GATC	11	33	0.024968	1
K09131	K09131	16	26	0.042825	1
K00963	UGP2, galU, galF	16	27	0.045093	1
KO	Gene ID	Smithii Subgroup 7 true	Smithii Subgroup 9 true	P value	Corrected P value
K01338	lon	6	0	0.023709	1

K03539	RPP1, RPP30	16	8	0.024176	1
K03537	POP5	16	8	0.024176	1
K19427	epsJ	8	1	0.038955	1
K06921	K06921	3	7	0.049668	1
KO	Gene ID	Smithii Subgroup 8 true	Smithii Subgroup 9 true	P value	Corrected P value
K02927	RP-L40e, RPL40	35	9	0.013568	1
K03539	RPP1, RPP30	33	8	0.030097	1
K03537	POP5	33	8	0.030097	1
K06921	K06921	8	7	0.034055	1
K19427	epsJ	15	1	0.037437	1
K07016	csm1, cas10	22	3	0.042385	1
K02013	ABC.FEV.A	13	9	0.042385	1
K09721	K09721	34	9	0.045945	1
K19075	cst2, cas7	1	3	0.045945	1
K02189	cbiG	34	9	0.045945	1
K09002	csm3	21	3	0.048991	1
K19138	csm2	21	3	0.048991	1
KO	Gene ID	Smithii Subgroup 2 true	Smithii Subgroup 10 true	P value	Corrected P value
K02016	ABC.FEV.S	2	10	0.002131	1
K02015	ABC.FEV.P	2	10	0.002131	1
K19115	csh2	1	7	0.006688	1
K02013	ABC.FEV.A	13	18	0.010736	1
K07487	K07487	2	8	0.011813	1
K19116	cas5h	1	6	0.016586	1
K19114	csh1	1	6	0.016586	1
K01953	asnB, ASNS	37	22	0.025098	1
K01533	copB	37	22	0.025098	1
K12410	npdA	30	26	0.035194	1
KO	Gene ID	Smithii Subgroup 3 true	Smithii Subgroup 10 true	P value	Corrected P value
K03658	helD	20	12	4.88E-05	0.055904429
K02864	RP-L10, MRPL10, rplJ	14	26	0.004138	0.947604518
K02867	RP-L11, MRPL11, rplK	14	26	0.004138	0.947604518
K02601	nusG	14	26	0.004138	0.947604518
K02863	RP-L1, MRPL1, rplA	14	26	0.004138	0.947604518
K07342	SEC61G, SSS1, secE	12	25	0.005961	1
K04096	smf	11	4	0.009852	1
K01155	E3.1.21.4	1	10	0.013007	1
KO	Gene ID	Smithii Subgroup 4 true	Smithii Subgroup 10 true	P value	Corrected P value
K00096	araM, egsA	32	21	0.014356	1
K00558	DNMT1, dcm	27	14	0.019218	1
K13812	fae-hps	32	22	0.035237	1
K19116	cas5h	1	6	0.037885	1
KO	Gene ID	Smithii Subgroup 5 true	Smithii Subgroup 10 true	P value	Corrected P value
K03658	helD	42	12	0.000254	0.30650784

K00002	AKR1A1, adh	32	25	0.003556	1
K00096	araM, egsA	48	21	0.004083	1
K19090	cas5t	11	0	0.006312	1
K19075	cst2, cas7	11	0	0.006312	1
K19088	cst1, cas8a	11	0	0.006312	1
K19114	csh1	1	6	0.006506	1
K19115	csh2	2	7	0.00742	1
K01874	MARS, metG	48	22	0.012993	1
K14623	dinD	2	6	0.01943	1
K19116	cas5h	2	6	0.01943	1
K02035	ABC.PE.S	48	23	0.040109	1
K14941	cofC	48	23	0.040109	1
K01881	PARS, proS	48	23	0.040109	1
K00286	proC	48	23	0.040109	1
K00402	mcrG	41	26	0.047421	1
K03421	mcrC	41	26	0.047421	1
K01953	asnB, ASNS	47	22	0.048631	1
K13812	fae-hps	47	22	0.048631	1
K01533	copB	47	22	0.048631	1
KO	Gene ID	Smithii Subgroup 6 true	Smithii Subgroup 10 true	P value	Corrected P value
K19115	csh2	0	7	0.002836	1
K19114	csh1	0	6	0.007091	1
K00558	DNMT1, dcm	26	14	0.00871	1
K19139	csm4	18	7	0.016786	1
K19091	cas6	1	7	0.019113	1
K07487	K07487	2	8	0.033158	1
K13812	fae-hps	30	22	0.040704	1
K01533	copB	30	22	0.040704	1
K19116	cas5h	1	6	0.041396	1
K19003	mgdA	14	5	0.047311	1
KO	Gene ID	Smithii Subgroup 7 true	Smithii Subgroup 10 true	P value	Corrected P value
K03658	helD	14	12	0.009632	1
K07487	K07487	0	8	0.015865	1
K19139	csm4	10	7	0.029125	1
KO	Gene ID	Smithii Subgroup 8 true	Smithii Subgroup 10 true	P value	Corrected P value
K02016	ABC.FEV.S	3	10	0.009577	1
K02015	ABC.FEV.P	3	10	0.009577	1
K00096	araM, egsA	35	21	0.011057	1
K02013	ABC.FEV.A	13	18	0.019724	1
K19139	csm4	20	7	0.022056	1
K13812	fae-hps	35	22	0.028648	1
K02927	RP-L40e, RPL40	35	22	0.028648	1
K03658	helD	26	12	0.033924	1
K09131	K09131	26	25	0.033973	1
KO	Gene ID	Smithii Subgroup 9 true	Smithii Subgroup 10 true	P value	Corrected P value
K10716	kch, trkA, mthK, pch	6	3	0.016397	1

K19075	cst2, cas7	3	0	0.026079	1
K07487	K07487	0	8	0.038518	1
KO	Gene ID	Smithii Subgroup 10 true	Smithii Subgroup 11 true	P value	Corrected P value
K19954	adh1	0	16	0.000251	0.203086165
K03658	helD	12	39	0.00069	0.203086165
K01841	pepM	1	18	0.000693	0.203086165
K09459	E4.1.1.82	1	18	0.000693	0.203086165
K13812	fae-hps	22	45	0.015386	1
K02016	ABC.FEV.S	10	7	0.0432	1
K02015	ABC.FEV.P	10	7	0.0432	1
K00765	hisG	23	45	0.04549	1
K02035	ABC.PE.S	23	45	0.04549	1
K14941	cofC	23	45	0.04549	1
K01873	VARS, valS	23	45	0.04549	1
K01802	E5.2.1.8	23	45	0.04549	1
K01881	PARS, proS	23	45	0.04549	1
K00789	metK	23	45	0.04549	1
K00963	UGP2, galU, galF	25	35	0.04632	1
K09131	K09131	25	35	0.04632	1
KO	Gene ID	Smithii Subgroup 2 true	Smithii Subgroup 11 true	P value	Corrected P value
K01841	pepM	0	18	4.19E-06	0.002595164
K09459	E4.1.1.82	0	18	4.19E-06	0.002595164
K19954	adh1	0	16	2.59E-05	0.010704086
K06921	K06921	18	5	0.000205	0.063439131
K06188	aqpZ	0	7	0.014645	1
K01953	asnB, ASNS	37	39	0.0299	1
K02428	rdgB	33	45	0.03776	1
K03496	parA, soj	33	45	0.03776	1
KO	Gene ID	Smithii Subgroup 3 true	Smithii Subgroup 11 true	P value	Corrected P value
K02867	RP-L11, MRPL11, rplK	14	45	0.000469	0.139133626
K02601	nusG	14	45	0.000469	0.139133626
K02863	RP-L1, MRPL1, rplA	14	45	0.000469	0.139133626
K01841	pepM	0	18	0.000571	0.139133626
K09459	E4.1.1.82	0	18	0.000571	0.139133626
K07342	SEC61G, SSS1, secE	12	43	0.00074	0.15032409
K19954	adh1	0	16	0.001345	0.215125993
K04096	smf	11	6	0.001412	0.215125993
K03100	lepB	15	45	0.001877	0.254232341
K02864	RP-L10, MRPL10, rplJ	14	44	0.002617	0.318975408
K10761	THG1	1	14	0.025572	1
K15342	cas1	19	31	0.025572	1
K01928	murE	17	45	0.026099	1
KO	Gene ID	Smithii Subgroup 4 true	Smithii Subgroup 11 true	P value	Corrected P value
K01841	pepM	0	18	1.43E-05	0.008367822

K09459	E4.1.1.82	0	18	1.43E-05	0.008367822
K19954	adh1	0	16	8.33E-05	0.032532878
K03100	lepB	27	45	0.010192	1
K09729	K09729	32	39	0.038176	1
K19003	mgdA	10	5	0.040678	1
K03658	helD	21	39	0.049055	1
KO	Gene ID	Smithii Subgroup 5 true	Smithii Subgroup 11 true	P value	Corrected P value
K09459	E4.1.1.82	0	18	2.36E-07	0.000297339
K19954	adh1	0	16	1.7E-06	0.001071446
K01841	pepM	1	18	3.35E-06	0.001409788
K00002	AKR1A1, adh	32	40	0.013138	1
K02016	ABC.FEV.S	18	7	0.020366	1
K02015	ABC.FEV.P	18	7	0.020366	1
K19114	csh1	1	7	0.02721	1
K02435	gatC, GATC	47	38	0.02721	1
K04096	smf	16	6	0.028945	1
K19427	epsJ	7	16	0.029474	1
K00721	DPM1	35	41	0.031475	1
K02189	cbiG	37	42	0.041157	1
KO	Gene ID	Smithii Subgroup 6 true	Smithii Subgroup 11 true	P value	Corrected P value
K19954	adh1	0	16	9.33E-05	0.119330871
K19003	mgdA	14	5	0.000896	0.412324132
K01841	pepM	2	18	0.00129	0.412324132
K09459	E4.1.1.82	2	18	0.00129	0.412324132
K19091	cas6	1	14	0.00294	0.752093559
K01154	hsdS	24	20	0.003708	0.790476374
K03100	lepB	25	45	0.008257	1
K07016	csm1, cas10	19	14	0.008825	1
K14654	RIB7, arfC	30	36	0.009267	1
K19139	csm4	18	13	0.009272	1
K02016	ABC.FEV.S	13	7	0.015107	1
K02015	ABC.FEV.P	13	7	0.015107	1
K19138	csm2	19	15	0.017361	1
K03088	rpoE	4	0	0.022547	1
K04771	degP, htrA	4	0	0.022547	1
K00558	DNMT1, dcm	26	29	0.03717	1
K19115	csh2	0	7	0.037351	1
K19114	csh1	0	7	0.037351	1
K01001	ALG7	30	38	0.037351	1
KO	Gene ID	Smithii Subgroup 7 true	Smithii Subgroup 11 true	P value	Corrected P value
K01841	pepM	0	18	0.00147	0.830816794
K09459	E4.1.1.82	0	18	0.00147	0.830816794
K01338	lon	6	2	0.002871	0.985235362
K03100	lepB	12	45	0.003488	0.985235362
K19954	adh1	0	16	0.006098	1
K19003	mgdA	7	5	0.009222	1

K19175	dptH	5	2	0.010764	1
K19174	dptG	5	2	0.010764	1
K07016	csm1, cas10	11	14	0.016332	1
K19140	csm5	10	13	0.03339	1
K19139	csm4	10	13	0.03339	1
K03462	NAMPT	15	30	0.046985	1
KO	Gene ID	Smithii Subgroup 8 true	Smithii Subgroup 11 true	P value	Corrected P value
K01841	pepM	0	18	5.42E-06	0.003191305
K09459	E4.1.1.82	0	18	5.42E-06	0.003191305
K19954	adh1	0	16	2.96E-05	0.011606626
K07016	csm1, cas10	22	14	0.006542	1
K19139	csm4	20	13	0.013056	1
K06188	aqpZ	0	7	0.016402	1
K09002	csm3	21	15	0.023781	1
K19138	csm2	21	15	0.023781	1
K15342	cas1	32	31	0.025588	1
K07149	K07149	31	45	0.033106	1
K03100	lepB	31	45	0.033106	1
K19140	csm5	19	13	0.037621	1
K09951	cas2	24	20	0.042055	1
KO	Gene ID	Smithii Subgroup 9 true	Smithii Subgroup 11 true	P value	Corrected P value
K06921	K06921	7	5	0.001477	1
K01841	pepM	0	18	0.011294	1
K09459	E4.1.1.82	0	18	0.011294	1
K19954	adh1	0	16	0.013192	1
K03539	RPP1, RPP30	8	43	0.014512	1
K03537	POP5	8	42	0.029693	1
K02866	RP-L10e, RPL10	10	45	0.041353	1
K02121	ATPVE, ntpE, atpE	10	45	0.041353	1
K07483	K07483	2	0	0.041353	1
K07149	K07149	10	45	0.041353	1
K01480	speB	10	45	0.041353	1

Table 6.10. KEGG Orthology (KO) annotated gene counts differentially enriched in *M. smithii* subgroups. EnrichM (v0.4.9) was used to determined KOs enriched in 11 *M. smithii* subgroups. KO annotations with P values ≤0.05 were included. A single count is determined by a single enrichment.

Gene ID	Count	Percentage	Gene Description
cas1	18	3.488372093	CRISP-associated protein Cas1
ABC.FEV.S	17	3.294573643	iron complex transport system substrate-binding protein
ABC.FEV.P	16	3.100775194	iron complex transport system permease protein
helD	14	2.713178295	DNA helicase IV [EC:3.6.4.12]
mgdA	13	2.519379845	1,2-diacylglycerol 3-beta-glucosyltransferase [EC:2.4.1.336]
csm4	12	2.325581395	CRISPR-associated protein Csm4
K06921	12	2.325581395	uncharacterized protein
csh1	11	2.131782946	CRISPR-associated protein Csh1
adh1	10	1.937984496	alcohol dehydrogenase [EC:1.1.1.-]
AKR1A1, adh	10	1.937984496	alcohol dehydrogenase (NADP+) [EC:1.1.1.2]
cas5h	10	1.937984496	CRISPR-associated protein Cas5h
csh2	10	1.937984496	CRISPR-associated protein Csh2
E4.1.1.82	10	1.937984496	phosphonopyruvate decarboxylase [EC:4.1.1.82]
pepM	10	1.937984496	phosphoenolpyruvate phosphomutase [EC:5.4.2.9]
THG1	10	1.937984496	tRNA(His) guanylyltransferase [EC:2.7.7.79]
hsdS	9	1.744186047	type I restriction enzyme, S subunit [EC:3.1.21.3]
SEC61G, SSS1, secE	9	1.744186047	protein transport protein SEC61 subunit gamma and related proteins
ABC.FEV.A	8	1.550387597	iron complex transport system ATP-binding protein [EC:3.6.3.34]
csm1, cas10	8	1.550387597	CRISPR-associated protein Csm1
epsJ	8	1.550387597	glycosyltransferase EpsJ [EC:2.4.-.-]
nusG	8	1.550387597	transcriptional antiterminator NusG
RP-L10, MRPL10, rplJ	8	1.550387597	large subunit ribosomal protein L10
RP-L11, MRPL11, rplK	8	1.550387597	large subunit ribosomal protein L11
smf	8	1.550387597	DNA processing protein
cas6	7	1.356589147	CRISPR-associated endoribonuclease Cas6 [EC:3.1.-.-]
cbiG	7	1.356589147	cobalt-precorrin 5A hydrolase [EC:3.7.1.12]
DPM1	7	1.356589147	dolichol-phosphate mannosyltransferase [EC:2.4.1.83]
dptH	7	1.356589147	DNA phosphorothioation-dependent restriction protein DptH
lepB	7	1.356589147	signal peptidase I [EC:3.4.21.89]
RP-L1, MRPL1, rplA	7	1.356589147	large subunit ribosomal protein L1
ACSF2	6	1.162790698	fatty-acyl-CoA synthase [EC:6.2.1.-]
aqpZ	6	1.162790698	aquaporin Z
csm2	6	1.162790698	CRISPR-associated protein Csm2
dptG	6	1.162790698	DNA phosphorothioation-dependent restriction protein DptG
fae-hps	6	1.162790698	bifunctional enzyme Fae/Hps [EC:4.2.1.147 4.1.2.43]
cst2, cas7	5	0.968992248	CRISPR-associated protein Cst2
epsH	5	0.968992248	glycosyltransferase EpsH [EC:2.4.-.-]
gatC, GATC	5	0.968992248	aspartyl-tRNA(Asn)/glutamyl-tRNA(Gln) amidotransferase subunit C [EC:6.3.5.6 6.3.5.7]
K07075	5	0.968992248	uncharacterized protein
K07487	5	0.968992248	transposase

K09131	5	0.968992248	uncharacterized protein
K09721	5	0.968992248	uncharacterized protein
araM, egsA	4	0.775193798	glycerol-1-phosphate dehydrogenase [NAD(P)+] [EC:1.1.1.261]
csm3	4	0.775193798	CRISPR-associated protein Csm3
csm5	4	0.775193798	CRISPR-associated protein Csm5
lon	4	0.775193798	ATP-dependent Lon protease [EC:3.4.21.53]
mcrG	4	0.775193798	methyl-coenzyme M reductase gamma subunit [EC:2.8.4.1]
RP-L40e, RPL40	4	0.775193798	large subunit ribosomal protein L40e
UGP2, galU, galF	4	0.775193798	UTP--glucose-1-phosphate uridylyltransferase [EC:2.7.7.9]
asnB, ASNS	3	0.581395349	asparagine synthase (glutamine-hydrolysing) [EC:6.3.5.4]
cas5t	3	0.581395349	CRISPR-associated protein Cas5t
copB	3	0.581395349	Cu2+-exporting ATPase [EC:3.6.3.4]
cst1, cas8a	3	0.581395349	CRISPR-associated protein Cst1
czcD, zitB	3	0.581395349	cobalt-zinc-cadmium efflux system protein
DNMT1, dcm	3	0.581395349	DNA (cytosine-5)-methyltransferase 1 [EC:2.1.1.37]
K07499	3	0.581395349	putative transposase
K09384	3	0.581395349	uncharacterized protein
mcrC	3	0.581395349	methyl-coenzyme M reductase subunit C
nadX, ASPDH	3	0.581395349	aspartate dehydrogenase [EC:1.4.1.21]
npdA	3	0.581395349	NAD-dependent deacetylase [EC:3.5.1.-]
POP5	3	0.581395349	ribonuclease P/MRP protein subunit POP5 [EC:3.1.26.5]
rpoE	3	0.581395349	RNA polymerase sigma-70 factor, ECF subfamily
RPP1, RPP30	3	0.581395349	ribonuclease P/MRP protein subunit RPP1 [EC:3.1.26.5]
trpE	3	0.581395349	anthranilate synthase component I [EC:4.1.3.27]
vsr	3	0.581395349	DNA mismatch endonuclease, patch repair protein [EC:3.1.-.-]
ABC.PE.S	2	0.387596899	peptide/nickel transport system substrate-binding protein
cbiB, cobD	2	0.387596899	adenosylcobinamide-phosphate synthase [EC:6.3.1.10]
cofC	2	0.387596899	2-phospho-L-lactate guanylyltransferase [EC:2.7.7.68]
cvrA, nhaP2	2	0.387596899	cell volume regulation protein A
degP, htrA	2	0.387596899	serine protease Do [EC:3.4.21.107]
E3.1.21.4	2	0.387596899	type II restriction enzyme [EC:3.1.21.4]
hsdM	2	0.387596899	type I restriction enzyme M protein [EC:2.1.1.72]
K06936	2	0.387596899	uncharacterized protein
K07068	2	0.387596899	uncharacterized protein
K07149	2	0.387596899	uncharacterized protein
K09729	2	0.387596899	uncharacterized protein
murE	2	0.387596899	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase [EC:6.3.2.13]
parA, soj	2	0.387596899	chromosome partitioning protein
PARS, proS	2	0.387596899	prolyl-tRNA synthetase [EC:6.1.1.15]
rdgB	2	0.387596899	XTP/dITP diphosphohydrolase [EC:3.6.1.66]
res	2	0.387596899	type III restriction enzyme [EC:3.1.21.5]
rfbC	2	0.387596899	O-antigen biosynthesis protein [EC:2.4.1.-]
sucD	2	0.387596899	succinyl-CoA synthetase alpha subunit [EC:6.2.1.5]
AARS,AG21:AG162 alaS	1	0.19379845	alanyl-tRNA synthetase [EC:6.1.1.7]
ALG7	1	0.19379845	UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminephosphotransferase [EC:2.7.8.15]
ATPVE, ntpE, atpE	1	0.19379845	V/A-type H+/Na+-transporting ATPase subunit E

ATPVK, ntpK, atpK	1	0.19379845	V/A-type H+/Na+-transporting ATPase subunit K
cas2	1	0.19379845	CRISPR-associated protein Cas2
dinD	1	0.19379845	DNA-damage-inducible protein D
dkgA	1	0.19379845	2,5-diketo-D-gluconate reductase A [EC:1.1.1.346]
E2.7.7.24, rfbA, rffH	1	0.19379845	glucose-1-phosphate thymidylyltransferase [EC:2.7.7.24]
E4.2.1.46, rfbB, rffG	1	0.19379845	dTDP-glucose 4,6-dehydratase [EC:4.2.1.46]
E5.2.1.8	1	0.19379845	peptidylprolyl isomerase [EC:5.2.1.8]
hisG	1	0.19379845	ATP phosphoribosyltransferase [EC:2.4.2.17]
hom	1	0.19379845	homoserine dehydrogenase [EC:1.1.1.3]
K07105	1	0.19379845	uncharacterized protein
K07483	1	0.19379845	transposase
K07579	1	0.19379845	putative methylase
K07744	1	0.19379845	transcriptional regulator
K09153	1	0.19379845	small membrane protein
K11646	1	0.19379845	3-dehydroquinate synthase II [EC:1.4.1.24]
kch, trkA, mthK, pch	1	0.19379845	voltage-gated potassium channel
MARS, metG	1	0.19379845	methionyl-tRNA synthetase [EC:6.1.1.10]
mcrA	1	0.19379845	methyl-coenzyme M reductase alpha subunit [EC:2.8.4.1]
mcrD	1	0.19379845	methyl-coenzyme M reductase subunit D
metK	1	0.19379845	S-adenosylmethionine synthetase [EC:2.5.1.6]
mscS	1	0.19379845	small conductance mechanosensitive channel
NAMPT	1	0.19379845	nicotinamide phosphoribosyltransferase [EC:2.4.2.12]
panE, apbA	1	0.19379845	2-dehydropantoate 2-reductase [EC:1.1.1.169]
proC	1	0.19379845	pyrroline-5-carboxylate reductase [EC:1.5.1.2]
recG	1	0.19379845	ATP-dependent DNA helicase RecG [EC:3.6.4.12]
rfbC, rmlC	1	0.19379845	dTDP-4-dehydrorhamnose 3,5-epimerase [EC:5.1.3.13]
RIB7, arfC	1	0.19379845	2,5-diamino-6-(ribosylamino)-4(3H)-pyrimidinone 5'-phosphate reductase [EC:1.1.1.302]
RP-L10e, RPL10	1	0.19379845	large subunit ribosomal protein L10e
speB	1	0.19379845	agmatinase [EC:3.5.3.11]
tet	1	0.19379845	tetrahedral aminopeptidase [EC:3.4.11.-]
trpA	1	0.19379845	tryptophan synthase alpha chain [EC:4.2.1.20]
trpC	1	0.19379845	indole-3-glycerol phosphate synthase [EC:4.1.1.48]
trpD	1	0.19379845	anthranilate phosphoribosyltransferase [EC:2.4.2.18]
trpG	1	0.19379845	anthranilate synthase component II [EC:4.1.3.27]
uvrA	1	0.19379845	excinuclease ABC subunit A
VARS, valS	1	0.19379845	valyl-tRNA synthetase [EC:6.1.1.9]
xseA	1	0.19379845	exodeoxyribonuclease VII large subunit [EC:3.1.11.6]

Table 6.11. Methanogen abundance in marsupial samples by amplicon-based sequencing. Methanogen sequences were recovered as per 4.2.2. Methanogenic archaea attributed to g__Methanobrevibacter;s__, g__Methanocorpusculum;s__, g__Methanocorpusculum;s__2 were detected across the samples.

No.	Sample ID	Marsupial	Species	g__Methanobrevibacter; s__	g__Methanocorpusculum; s__	g__Methanocorpusculum; s__2
204	S0487	Glider - Greater	*Petauroides volans*	0	0	0
205	S0488	Glider - Greater	*Petauroides volans*	0	0	0
213	S1398	Glider - Greater	*Petauroides volans*	0	0	0
215	S1400	Glider - Greater	*Petauroides volans*	0	0	0
219	S7502	Glider - Mahogany	*Petaurus gracilis*	0	0.024630542	0.007983693
218	S7501	Glider - Mahogany	*Petaurus gracilis*	0	0.002409923	0.000590667
208	S7500	Glider - Mahogany	*Petaurus gracilis*	0	0.022197589	0.006581682
207	S7499	Glider - Mahogany	*Petaurus gracilis*	0	0.008974973	0.00354742
193	S7496	Glider - Mahogany	*Petaurus gracilis*	0	0.000100338	1.82433E-05
187	S2020	Glider - Mahogany	*Petaurus gracilis*	0	0	9.48021E-05
192	S0481	Glider - Mahogany	*Petaurus gracilis*	0	0	0
167	S0464	Glider - Mahogany	*Petaurus gracilis*	1.94032E-05	0.002308976	0.001804494
166	S0463	Glider - Mahogany	*Petaurus gracilis*	0	0.000206827	0.008389986
194	S7497	Glider - Squirrel	*Petaurus norfolcensis*	0	0.010015869	0.003355967
214	S1399	Glider - Squirrel	*Petaurus norfolcensis*	0	0.000145353	0.007280848
211	S1396	Glider - Squirrel	*Petaurus norfolcensis*	0	0.000162494	0.003974851
209	S0489	Glider - Squirrel	*Petaurus norfolcensis*	0	0.000245187	0.007237559
206	S7498	Glider - Yellow-bellied	*Petaurus australis*	0	0	0
53	S7486	Kangaroo - Eastern Grey	*Macropus giganteus*	0.000425003	0	0.011506556
50	S7485	Kangaroo - Eastern Grey	*Macropus giganteus*	0.000693742	0	0.00205232
186	S2019	Kangaroo - Eastern Grey	*Macropus giganteus*	0.000166967	0	0
185	S0474	Kangaroo - Eastern Grey	*Macropus giganteus*	9.26827E-06	0	0.002233653
170	S0466	Kangaroo - Eastern Grey	*Macropus giganteus*	0.009661391	0	0.000214698
169	S0465	Kangaroo - Eastern Grey	*Macropus giganteus*	0.000674717	8.26185E-05	0.003483745
49	S0337	Kangaroo - Eastern Grey	*Macropus giganteus*	0	0	0.003215006
46	S0334	Kangaroo - Eastern Grey	*Macropus giganteus*	0.000685525	0	0.005110279
52	S0339	Kangaroo - Red	*Macropus rufus*	0.001227858	0	0.000374677
48	S0336	Kangaroo - Red	*Macropus rufus*	0.001713856	0	1.76081E-05
45	S0333	Kangaroo - Red	*Macropus rufus*	0.003834653	0	4.15305E-05
198	S0484	Koala	*Phascolarctos cinereus*	0.026363983	0	0
196	S0483	Koala	*Phascolarctos cinereus*	3.21038E-05	0	0
179	S0471	Koala	*Phascolarctos cinereus*	0	0	0
178	S0470	Koala	*Phascolarctos cinereus*	0	0	0
101	S0343	Koala	*Phascolarctos cinereus*	0	0	0
30	S0332	Koala	*Phascolarctos cinereus*	0	7.66812E-05	0
19	S0331	Koala	*Phascolarctos cinereus*	0	0	0
8	S0330	Koala	*Phascolarctos cinereus*	0	0	1.71494E-05
6	S0329	Koala	*Phascolarctos cinereus*	0	0	0

37	S0094	Koala	*Phascolarctos cinereus*	2.22762E-05	0	0
235	S7878	Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	0	0	1.13654E-05
234	S7877	Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	0	3.77986E-05	3.77986E-05
233	S2033	Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	0	0	0
232	S2032	Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	0	0	0
168	S7492	Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	0.00128999	0.000859993	0.000257998
203	S0486	Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	9.10631E-06	0	0.001784836
202	S0485	Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	0	0	0.000375537
56	S0340	Pandemelon - Red-legged	*Thylogale stigmatica*	8.53197E-05	5.33248E-06	0.001503759
231	S2031	Pandemelon - Red-legged	*Thylogale stigmatica*	0	0	0.001213903
230	S2030	Pandemelon - Red-legged	*Thylogale stigmatica*	1.09498E-05	0	0.000722686
229	S2029	Pandemelon - Red-legged	*Thylogale stigmatica*	0	0	0.000545408
77	S0095	Possum	*Phalangeridae*	0	0.000140697	1.27907E-05
160	S7488	Possum - Common brushtail	*Trichosurus vulpecula*	0	0	0
159	S0346	Possum - Common brushtail	*Trichosurus vulpecula*	0	0	0
212	S1397	Possum - Common brushtail	*Trichosurus vulpecula*	0	0	0
165	S7491	Possum - Common brushtail	*Trichosurus vulpecula*	0	0	0
164	S7490	Possum - Common brushtail	*Trichosurus vulpecula*	0	0	0
238	S2036	Possum - Common brushtail	*Trichosurus vulpecula*	0	0	0
223	S2023	Possum - Common brushtail	*Trichosurus vulpecula*	1.52228E-05	0	0
221	S1404	Possum - Common ringtail	*Pseudocheirus peregrinus*	0	0	0
217	S1402	Possum - Common ringtail	*Pseudocheirus peregrinus*	0	0	0
210	S0490	Possum - Common ringtail	*Pseudocheirus peregrinus*	0	0	0
195	S0482	Possum - Common ringtail	*Pseudocheirus peregrinus*	0.001075599	0	0
162	S7489	Possum - Common ringtail	*Pseudocheirus peregrinus*	0	0	0
158	S7487	Possum - Golden brushtail	*Trichosurus vulpecula*	1.63052E-05	0	0
154	S0345	Possum - Golden brushtail	*Trichosurus vulpecula*	1.05674E-05	5.28368E-06	0
237	S2035	Possum - Green ringtail	*Pseudochirops archeri*	0	0	0
236	S2034	Possum - Green ringtail	*Pseudochirops archeri*	0	0	0
228	S2028	Possum - Green ringtail	*Pseudochirops archeri*	0	0	0
227	S2027	Possum - Green ringtail	*Pseudochirops archeri*	0	0	0
226	S2026	Possum - Green ringtail	*Pseudochirops archeri*	0	0	0
225	S2025	Possum - Herbet River ringtail	*Pseudochirulus herbertensis*	0	0	0
224	S2024	Possum - Herbet River ringtail	*Pseudochirulus herbertensis*	0	0	0
161	S0347	Possum - Mountain brushtail	*Trichosurus cunninghami*	0	0	0
222	S1405	Possum - Short-eared	*Trichosurus caninus*	0	0	0
220	S1403	Possum - Short-eared	*Trichosurus caninus*	0	0	1.73575E-05
216	S1401	Possum - Short-eared	*Trichosurus caninus*	0	0	0
181	S0473	Possum - Striped	*Dactylopsila trivirgata*	0	0.001904481	0.000536622
180	S0472	Possum - Striped	*Dactylopsila trivirgata*	0	0.000126048	0
177	S7874	Wallaby - Agile	*Macropus agilis*	0	0	0.000471032
176	S7494	Wallaby - Agile	*Macropus agilis*	0.000376419	0	0.000376419
175	S7493	Wallaby - Agile	*Macropus agilis*	0.000381882	0	0.000742548

174	S0469	Wallaby - Northern nailtail	*Onychogalea unguifera*	0.001086433	0	0
173	S0468	Wallaby - Northern nailtail	*Onychogalea unguifera*	0.00033447	0	0
172	S0467	Wallaby - Northern nailtail	*Onychogalea unguifera*	0.001327573	0	0
183	S7876	Wallaby - Parma	*Macropus parma*	0.000105119	0	0
182	S7875	Wallaby - Parma	*Macropus parma*	0	0	0.000799087
184	S7495	Wallaby - Parma	*Macropus parma*	0.000726662	0	5.67704E-05
63	S0342	Wallaby - Red-necked	*Macropus rufogriseus*	0.002391203	0	0.001024801
51	S0338	Wallaby - Red-necked	*Macropus rufogriseus*	0	0	2.20325E-05
58	S0341	Wallaby - Swamp	*Wallabia bicolor*	0.000312562	0	0.000252618
47	S0335	Wallaby - Swamp	*Wallabia bicolor*	4.8479E-05	0	8.88781E-05
191	S0480	Wombat - Common	*Vombatus ursinus*	0.006774748	8.73035E-06	0.000192068
190	S0479	Wombat - Common	*Vombatus ursinus*	0.030072934	6.66067E-05	0.001154516
153	S0344	Wombat - Common	*Vombatus ursinus*	0	0.000387463	0.005319251
35	S0093	Wombat - Common	*Vombatus ursinus*	0	0.000228292	0.002511215
189	S2022	Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	0.002144561	0	0.000175066
188	S2021	Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	0.055230622	0	0
106	S0092	Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	0	0.004093671	0.038618608
57	S0091	Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	0	0.000715662	0.007129094
44	S0042	Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	0	0.003318796	0.041441907
36	S0041	Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	0	0.003493517	0.038078237
26	S0009	Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	0	0.005353884	0.04859055

Table 6.12. Average methanogen abundance detected in marsupial faecal samples with amplicon sequencing. Methanogen sequences were detected as per Section 4.2.2 and values averages of those shown in Table 6.11. *Pseudochirulus herbertensis* (n=2), *Dactylopsila trivirgata* (n=2), *Dendrolagus lumholtzi* (n=7), *Lasiorhinus latifrons* (n=7), *Macropus agilis* (n=3), *Macropus giganteus* (n=8), *Macropus rufogriseus* (n=2), *Macropus parma* (n=3), *Macropus rufus* (n=3), *Onychogalea unguifera* (n=3), *Petauroides Volans* (n=4), *Petaurus gracilis* (n=9), *Petaurus norfolcensis* (n=4), *Petaurus australis* (n=1), *Phascolarctos cinereus* (n=10), *Pseudocheirus peregrinus* (n=5), *Pseudochirops archeri* (n=5), *Thylogale stigmatica* (n=4), *Trichosurus vulpecula* (n=9), *Trichosurus cunninghami* (n=1), *Trichosurus caninus* (n=3), *Vombatus ursinus* (n=4) and *Wallabia bicolor* (n=2).

Marsupial	**Species**	**Methanobrev.**	**Methanocorp.**	**Methanocorp. 2**	**Methanobrev. std**	**Methanocorp. std**	**Methanocorp. 2 std**
Possum - Herbet River ringtail	*Pseudochirulus herbertensis*	0	0	0	0	0	0
Possum - Striped	*Dactylopsila trivirgata*	0	0	0	0	0	0
Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	0	0.008330481	0.002688073	0	0.010801	0.003401
Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	2.77188E-06	0.001869244	0.004577672	7.3337E-06	0.003682	0.003147
Wallaby - Agile	*Macropus agilis*	0.00042857	0	0.004519625	0.000263406	0	0.006137
Kangaroo - Eastern Grey	*Macropus giganteus*	0.002225909	1.03273E-05	0.0018364	0.003246534	2.92E-05	0.001956
Wallaby - Red-necked	*Macropus rufogriseus*	0.013198043	0	0	0.018619451	0	0
Wallaby - Parma	*Macropus parma*	0	0	0	0	0	0
Wallaby - Northern nailtail	*Onychogalea unguifera*	0	0	5.71647E-06	0	0	9.9E-06
Kangaroo - Red	*Macropus rufus*	7.42539E-06	1.25995E-05	1.6388E-05	1.28612E-05	2.18E-05	1.94E-05
Glider - Greater	*Petauroides volans*	0.000324774	0.000214998	0.000510709	0.000643491	0.00043	0.000858
Glider - Mahogany	*Petaurus gracilis*	1.06966E-05	1.62255E-05	0.000486009	2.8217E-05	4.67E-05	0.000567
Glider - Squirrel	*Petaurus norfolcensis*	3.8057E-06	0	0	7.61139E-06	0	0
Glider - Yellow-bellied	*Petaurus australis*	0	0	0	0	0	0
Koala	*Phascolarctos cinereus*	0.000110247	5.28368E-07	0	0.000339239	1.67E-06	0
Possum - Common ringtail	*Pseudocheirus peregrinus*	0	0	0	0	0	0
Possum - Green ringtail	*Pseudochirops archeri*	0	0.000406106	0.000205002	0	0	0
Pandemelon - Red-legged	*Thylogale stigmatica*	0.000544801	0	0.000279742	0.000361709	0	0.000356
Possum - Common brushtail	*Trichosurus vulpecula*	0.000694731	0	0.000376132	0.000888681	0	0.000427
Possum - Golden brushtail	*Trichosurus vulpecula*	0.003411614	4.36517E-06	0.000140473	0.004756191	6.17E-06	7.3E-05
Possum - Mountain brushtail	*Trichosurus cunninghami*	0.030072934	6.66067E-05	0.001154516	0	0	0
Possum - Short-eared	*Trichosurus caninus*	0.000714854	0.000205252	0.002668511	0.001238163	0.000195	0.002576
Wombat - Common	*Vombatus ursinus*	0.013807656	0.002032032	0.021797402	0.027615311	0.001981	0.021285
Wallaby - Swamp	*Wallabia bicolor*	0	0.0044237	0.043334393	0	0.001315	0.007433

Table 6.13. Methanogen abundance in marsupial faecal samples by metagenomic sequencing (MGS). Methanogens were identified using CoverM (v0.6.0) with the GTDB-tk database and MAGs recovered from the marsupial sequences used for reference. Only reference genomes that were detected were included. S0092_shn_wombat_3 was classified to the genus *Methanocorpusculum*, S2021_shn_wombat_3 was classified to the order *Methanomassiliicoccales* and S1403_short_earred_possum_11 was classified to the *Methanomethylophilaceae* genus UBA71.

Sample ID	Marsupial	Species	Methanocorpusculum sp. Phil4	S0092_shn_wombat_3	S2021_shn_wombat_1	S1403_short_earred_possum_11	M. smithii	M. alvus	M. flavescens	M. thermophila
S0487	Glider - Greater	*Petauroides volans*	0	0	0	0	0	0	0	0
S0488	Glider - Greater	*Petauroides volans*	0	0	0	0	0	0	0	0
S1398	Glider - Greater	*Petauroides volans*	0	0	0	0	0	0	0	0
S1400	Glider - Greater	*Petauroides volans*	0	0	0	0	0	0	0	0
S0463	Glider - Mahogany	*Petaurus gracilis*	0	0	0	0	0	0	0	0
S0464	Glider - Mahogany	*Petaurus gracilis*	0	0	0	0	0	0	0	0
S0481	Glider - Mahogany	*Petaurus gracilis*	0	0	0	0	0	0	0	0
S2020	Glider - Mahogany	*Petaurus gracilis*	0	0	0	0	0	0	0	0
S7496	Glider - Mahogany	*Petaurus gracilis*	0	0	0	0	0	0	0	0
S7499	Glider - Mahogany	*Petaurus gracilis*	0	0	0	0	0	0	0	0
S7500	Glider - Mahogany	*Petaurus gracilis*	0	0	0	0	0	0	0	0
S7501	Glider - Mahogany	*Petaurus gracilis*	0	0	0	0	0	0	0	0
S7502	Glider - Mahogany	*Petaurus gracilis*	0	0	0	0	0	0	0	0
S0489	Glider - Squirrel	*Petaurus norfolcensis*	0	0	0	0	0	0	0	0
S1396	Glider - Squirrel	*Petaurus norfolcensis*	0	0	0	0	0	0	0	0
S1399	Glider - Squirrel	*Petaurus norfolcensis*	0	0	0	0	0	0	0	0
S7497	Glider - Squirrel	*Petaurus norfolcensis*	0	0	0	0	0	0	0	0
S7498	Glider - Yellow-bellied	*Petaurus australis*	0	0	0	0	0	0	0	0
S0334	Kangaroo - Eastern Grey	*Macropus giganteus*	0	0	0	0	0	0	0	0
S0337	Kangaroo - Eastern Grey	*Macropus giganteus*	0	0	0	0	0	0	0	0
S2019	Kangaroo - Eastern Grey	*Macropus giganteus*	0	0	0	0	0	0	0	0
S7485	Kangaroo - Eastern Grey	*Macropus giganteus*	0	0	0	0	0	0	0	0
S7486	Kangaroo - Eastern Grey	*Macropus giganteus*	0	0	0	0	0	0	0	0

S0333	Kangaroo - Red	*Macropus rufus*	0	0	0	0	0	0	0	0
S0336	Kangaroo - Red	*Macropus rufus*	0	0	0	0	0	0	0	0
S0339	Kangaroo - Red	*Macropus rufus*	0	0	0	0	0	0	0	0
S0332	Koala	*Phascolarctos cinereus*	0	0	0	0	0	0	0	0
S0343	Koala	*Phascolarctos cinereus*	0	0	0	0	0	0	0	0
S0470	Koala	*Phascolarctos cinereus*	0	0	0	0	0	0	0	0
S0471	Koala	*Phascolarctos cinereus*	0	0	0	0	0	0	0	0
S0483	Koala	*Phascolarctos cinereus*	0	0	0	0	0	0	0	0
S0484	Koala	*Phascolarctos cinereus*	0	0	0.35541478	0	0	0	0	0
S0485	Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	0	0	0	0	0	0	0	0
S0486	Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	0	0	0	0	0	0	0	0
S2032	Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	0	0	0	0	0	0	0	0
S2033	Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	0	0	0	0	0	0	0	0
S7492	Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	0	0	0	0	0	0	0	0
S7877	Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	0	0	0	0	0	0	0	0
S7878	Lumholtz's tree kangaroo	*Dendrolagus lumholtzi*	0	0	0	0	0	0	0	0
S0340	Pandemelon - Red-legged	*Thylogale stigmatica*	0	0	0	0	0	0	0	0
S2029	Pandemelon - Red-legged	*Thylogale stigmatica*	0	0	0	0	0	0	0	0
S2030	Pandemelon - Red-legged	*Thylogale stigmatica*	0	0	0	0	0	0	0	0
S2031	Pandemelon - Red-legged	*Thylogale stigmatica*	0	0	0	0	0	0	0	0
S0095	Possum - ?		0	0	0	0.3508985	0	0	0	0
S0346	Possum - Common brushtail	*Trichosurus vulpecula*	0	0	0	0.105877355	0	0	0	0
S1397	Possum - Common brushtail	*Trichosurus vulpecula*	0	0	0	0	0	0	0	0
S2023	Possum - Common brushtail	*Trichosurus vulpecula*	0	0	0	0.014015016	0	0	0	0
S2036	Possum - Common brushtail	*Trichosurus vulpecula*	0	0	0	0.013341444	0	0	0	0
S7488	Possum - Common brushtail	*Trichosurus vulpecula*	0	0	0	0	0	0	0	0
S7490	Possum - Common brushtail	*Trichosurus vulpecula*	0	0	0	0	0	0	0	0
S7491	Possum - Common brushtail	*Trichosurus vulpecula*	0	0	0	0	0	0	0	0

S0482	Possum - Common ringtail	*Pseudocheirus peregrinus*	0	0	0	0	0	0	0	0
S0490	Possum - Common ringtail	*Pseudocheirus peregrinus*	0	0	0	0	0	0	0	0
S1402	Possum - Common ringtail	*Pseudocheirus peregrinus*	0	0	0	0	0	0	0	0
S1404	Possum - Common ringtail	*Pseudocheirus peregrinus*	0	0	0	0	0	0	0	0
S7489	Possum - Common ringtail	*Pseudocheirus peregrinus*	0	0	0	0	0	0	0	0
S0345	Possum - Golden brushtail	*Trichosurus vulpecula*	0	0	0	0.047106843	0	0	0	0
S7487	Possum - Golden brushtail	*Trichosurus vulpecula*	0	0	0	0.067025326	0	0	0	0
S2026	Possum - Green ringtail	*Pseudochirops archeri*	0	0	0	0	0	0	0	0
S2027	Possum - Green ringtail	*Pseudochirops archeri*	0	0	0	0	0	0	0	0
S2028	Possum - Green ringtail	*Pseudochirops archeri*	0	0	0	0	0	0	0	0
S2034	Possum - Green ringtail	*Pseudochirops archeri*	0	0	0	0	0	0	0	0
S2035	Possum - Green ringtail	*Pseudochirops archeri*	0	0	0	0	0	0	0	0
S2024	Possum - Herbet River ringtail	*Pseudochirulus herbertensis*	0	0	0	0	0	0	0	0
S2025	Possum - Herbet River ringtail	*Pseudochirulus herbertensis*	0	0	0	0	0	0	0	0
S0347	Possum - Mountain brushtail	*Trichosurus cunninghami*	0	0	0	0.29838794	0	0	0	0
S1401	Possum - Short-eared	*Trichosurus caninus*	0	0	0	0.4666876	0	0	0	0
S1403	Possum - Short-eared	*Trichosurus caninus*	0	0	0	0.55185306	0	0	0	0
S1405	Possum - Short-eared	*Trichosurus caninus*	0	0	0	0.28712812	0	0	0	0
S0472	Possum - Striped	*Dactylopsila trivirgata*	0	0	0	0	0	0	0	0
S0473	Possum - Striped	*Dactylopsila trivirgata*	0	0	0	0	0	0	0	0
S7493	Wallaby - Agile	*Macropus agilis*	0	0	0	0	0	0	0	0
S7494	Wallaby - Agile	*Macropus agilis*	0	0	0	0	0	0	0	0
S7874	Wallaby - Agile	*Macropus agilis*	0	0	0	0	0	0	0	0
S0467	Wallaby - Northern nailtail	*Onychogalea unguifera*	0	0	0	0	0	0	0	0
S0468	Wallaby - Northern nailtail	*Onychogalea unguifera*	0	0	0	0	0	0	0	0
S0469	Wallaby - Northern nailtail	*Onychogalea unguifera*	0	0	0	0	0	0	0	0
S7495	Wallaby - Parma	*Macropus parma*	0	0	0	0	0	0	0	0
S7875	Wallaby - Parma	*Macropus parma*	0	0	0	0	0	0	0	0
S7876	Wallaby - Parma	*Macropus parma*	0	0	0	0	0	0	0	0

S0338	Wallaby - Red-necked	*Macropus rufogriseus*	0	0	0	0	0	0	0	0
S0342	Wallaby - Red-necked	*Macropus rufogriseus*	0	0	0	0	0	0	0	0
S0335	Wallaby - Swamp	*Wallabia bicolor*	0	0	0	0	0	0	0	0
S0341	Wallaby - Swamp	*Wallabia bicolor*	0	0	0	0	0	0	0	0
S0093	Wombat - Common	*Vombatus ursinus*	0.10022785	0.1053466	0.027311595	0	0	0	0	0
S0344	Wombat - Common	*Vombatus ursinus*	0.12889858	0.13388494	0.057654716	0	0	0.011559804	0	0
S0479	Wombat - Common	*Vombatus ursinus*	0	0	0.03293061	0	0	0	0	0
S0480	Wombat - Common	*Vombatus ursinus*	0	0	0.19016322	0	0	0	0	0
S0009	Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	1.9766595	1.8827337	0.39421222	0	0	0	0	0
S0041	Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	1.3144999	1.1237111	0.2995605	0	0	0	0.029794335	0.012568471
S0042	Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	1.8609165	1.7421637	0.33300927	0	0	0	0	0
S0091	Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	0.77973366	0.6460838	0.29262355	0	0	0	0	0
S0092	Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	0.9779114	1.0779903	0.32915112	0	0	0	0.012497839	0
S2021	Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	0	0	0.88478315	0	0.012548808	0.018857067	0	0
S2022	Wombat - Southern hairy-nosed	*Lasiorhinus latifrons*	0	0	0.6051684	0	0.12801984	0.06017124	0	0

Table 6.14. Average methanogen abundance detected in marsupial faecal samples with metagenomic sequencing (MGS). Methanogen sequences were detected as per Section 4.2.2 and values represent average of those shown in 6.13. *Pseudochirulus herbertensis* (n=2), *Dactylopsila trivirgata* (n=2), *Dendrolagus lumholtzi* (n=7), *Lasiorhinus latifrons* (n=7), *Macropus agilis* (n=3), *Macropus giganteus* (n=5), *Macropus rufogriseus* (n=2), *Macropus parma* (n=3), *Macropus rufus* (n=3), *Onychogalea unguifera* (n=3), *Petauroides Volans* (n=4), *Petaurus gracilis* (n=9), *Petaurus norfolcensis* (n=4), *Petaurus australis* (n=1), *Phascolarctos cinereus* (n=6), *Pseudocheirus peregrinus* (n=5), *Pseudochirops archeri* (n=5), *Thylogale stigmatica* (n=4), *Trichosurus vulpecula* (n=9), *Trichosurus cunninghami* (n=1), *Trichosurus caninus* (n=3), *Vombatus ursinus* (n=4) and *Wallabia bicolor* (n=2).

Average abundance									
Species	Animal	Methanocorpusculum sp. Phil4	S0092_shn_wombat_3	S2021_shn_wombat_1	S1403_short_earred_possum_11	M. smithii	M. alvus	M. flavescens	M. thermophila
Dactylopsila trivirgata	Possum - Striped	0	0	0	0	0	0	0	0
Dendrolagus lumholtzi	Lumholtz's tree kangaroo	0	0	0	0	0	0	0	0
Lasiorhinus latifrons	Wombat - Southern hairy-nosed	0.987102994	0.924668943	0.448358316	0	0.020081235	0.011289758	0.006041739	0.001795496
Macropus agilis	Wallaby - Agile	0	0	0	0	0	0	0	0
Macropus giganteus	Kangaroo - Eastern Grey	0	0	0	0	0	0	0	0
Macropus parma	Wallaby - Parma	0	0	0	0	0	0	0	0
Macropus rufogriseus	Wallaby - Red-necked	0	0	0	0	0	0	0	0
Macropus rufus	Kangaroo - Red	0	0	0	0	0	0	0	0
Onychogalea unguifera	Wallaby - Northern nailtail	0	0	0	0	0	0	0	0
Petauroides volans	Glider - Greater	0	0	0	0	0	0	0	0
Petaurus australis	Glider - Yellow-bellied	0	0	0	0	0	0	0	0
Petaurus gracilis	Glider - Mahogany	0	0	0	0	0	0	0	0
Petaurus norfolcensis	Glider - Squirrel	0	0	0	0	0	0	0	0
Phascolarctos cinereus	Koala	0	0	0.059235797	0	0	0	0	0
Pseudocheirus peregrinus	Possum - Common ringtail	0	0	0	0	0	0	0	0
Pseudochirops archeri	Possum - Green ringtail	0	0	0	0	0	0	0	0
Pseudochirulus herbertensis	Possum - Herbet River ringtail	0	0	0	0	0	0	0	0
Thylogale stigmatica	Pandemelon - Red-legged	0	0	0	0	0	0	0	0
Trichosurus caninus	Possum - Short-eared	0	0	0	0.435222927	0	0	0	0
Trichosurus cunninghami	Possum - Mountain brushtail	0	0	0	0.29838794	0	0	0	0
Trichosurus vulpecula	Possum - Common brushtail	0	0	0	0.027485109	0	0	0	0
Vombatus ursinus	Wombat - Common	0.057281608	0.059807885	0.077015035	0	0	0.002889951	0	0

Wallabia bicolor	Wallaby - Swamp	0	0	0	0	0	0	0	0
Standard deviation									
Species	Animal	Methanocorpusculum sp. Phil4	S0092_shn_wombat_3	S2021_shn_wombat_1	S1403_short_earred_possum_11	M. smithii	M. alvus	M. flavescens	M. thermophila
Dactylopsila trivirgata	Possum - Striped	0	0	0	0	0	0	0	0
Dendrolagus lumholtzi	Lumholtz's tree kangaroo	0	0	0	0	0	0	0	0
Lasiorhinus latifrons	Wombat - Southern hairy-nosed	0.800186647	0.756851589	0.220566582	0	0.047825655	0.022671409	0.011462839	0.004750436
Macropus agilis	Wallaby - Agile	0	0	0	0	0	0	0	0
Macropus giganteus	Kangaroo - Eastern Grey	0	0	0	0	0	0	0	0
Macropus parma	Wallaby - Parma	0	0	0	0	0	0	0	0
Macropus rufogriseus	Wallaby - Red-necked	0	0	0	0	0	0	0	0
Macropus rufus	Kangaroo - Red	0	0	0	0	0	0	0	0
Onychogalea unguifera	Wallaby - Northern nailtail	0	0	0	0	0	0	0	0
Petauroides volans	Glider - Greater	0	0	0	0	0	0	0	0
Petaurus australis	Glider - Yellow-bellied	-	-	-	-	-	-	-	-
Petaurus gracilis	Glider - Mahogany	0	0	0	0	0	0	0	0
Petaurus norfolcensis	Glider - Squirrel	0	0	0	0	0	0	0	0
Phascolarctos cinereus	Koala	0	0	0.145097476	0	0	0	0	0
Pseudocheirus peregrinus	Possum - Common ringtail	0	0	0	0	0	0	0	0
Pseudochirops archeri	Possum - Green ringtail	0	0	0	0	0	0	0	0
Pseudochirulus herbertensis	Possum - Herbet River ringtail	0	0	0	0	0	0	0	0
Thylogale stigmatica	Pandemelon - Red-legged	0	0	0	0	0	0	0	0
Trichosurus caninus	Possum - Short-eared	0	0	0	0.135138236	0	0	0	0
Trichosurus cunninghami	Possum - Mountain brushtail	-	-	-	-	-	-	-	-
Trichosurus vulpecula	Possum - Common brushtail	0	0	0	0.037909427	0	0	0	0
Vombatus ursinus	Wombat - Common	0.067170766	0.070036064	0.076575022	0	0	0.005779902	0	0
Wallabia bicolor	Wallaby - Swamp	0	0	0	0	0	0	0	0

Table 6.15. List of publicly available metagenomes used for the recovery of methanogen MAGs.

Author	BioProject	PMID	Animal/s	MAG Prefix	Samples Analysed	MAGs Recovered
-	PRJNA590977	-	Horse	PRJNA590977	5	5
-	PRJNA545600	-	Elephant	-	1	4
-	PRJNA545601	-	Elephant	-	1	1
-	PRJNA545606	-	Elephant	-	1	2
-	PRJNA293646	-	Chicken, Pig, bovine	-	13	5
(Campbell et al., 2020)	PRJNA539933	32203121	Chimpanzee	CampbellTP	159	7
(Doster et al., 2018)	PRJNA309291	30105011	Bovine	DosterE	60	6
(Salgado-Flores et al., 2019)	PRJNA450906	30856229	Ptarmigan	SalgadoFloresA	14	18
(Gibson et al., 2019)	PRJNA532626	31138833	Rhinoceros	GibsonKM	25	24
(Ilmberger et al., 2014)	PRJNA240141	25208077	Elephant	IlmbergerN	2	2
(Lee et al., 2014)	PRJEB1391	24108330	Bovine, Human, Pig, Chicken	LeeS	4	3
(Li et al., 2019)	PRJNA411766	31309006	Sperm Whale	LiC	1	3
(Lim et al., 2020)	PRJEB32496	-	Bovine, cattle	LimSK	77	41
(Rovira Sanz, 2017)	PRJNA379303	-	Bovine	SanzPR	94	47
(Chung et al., 2020)	PRJEB33889	31978162	Mouse	-	8	0
(Deusch et al., 2015)	PRJEB9357	26659594	Cat	-	88	0
(Donovan et al., 2020)	PRJNA449069	31919172	Vole	-	8	0
(Guo et al., 2019)	PRJNA356809, PRJNA358755	31635158	Panda, Bear, Pig	-	24	0
(He et al., 2018)	PRJNA413828, PRJNA477940	30509257	Tiger	-	9	0
(Zhang et al., 2019)	PRJNA492972	31337187	Elephant	-	3	0
(Qin et al., 2020)	PRJNA560663	32117147	Gazelle, Sheep	QinW	8	1
(C. Wang et al., 2019)	PRJEB31742	31848308	Pig	WangC	4	5
(Tan et al., 2017)	PRJNA389749	28848539	Pig	TanZ	11	1
(Svartström et al., 2017)	PRJEB12797	28731473	Moose	SvartstromO	7	2
(Gong et al., 2020)	PRJNA624740	33036549	Yak	GongG	5	1
-	PRJNA473026	-	Pig	-	1	0
-	PRJNA471970	-	Musk Deer	-	1	0
-	PRJNA471984	-	Yak	-	1	0
-	PRJNA471973	-	Masked palm civet	-	1	0
-	PRJNA471974	-	Chicken	-	1	0
(Cao et al., 2019)	PRJNA485657	30891061	Goat	CaoY	17	6
(Hou et al., 2016)	PRJNA340908	27876778	Chicken	HouQ	29	18
(Glendinning et al., 2021)	PRJEB34458	33479378	Red Deer, Reindeer, Cow, Sheep	GlendinningL	12	9
(Ghanbari et al., 2019)	PRJNA471402	30858509	Pig	GhanbariM	39	5

(Q. Chen et al., 2021)	PRJNA524932		Pig	ChenQ	20	13
(Chekabab et al., 2020)	PRJNA633402, PRJNA633385, PRJNA633399, PRJNA633392	33033582	Pig	ChekababS	20	7
-	PRJNA260105		Camel	-	1	0
-	PRJNA396259		Camel	-	1	0
(Chen et al., 2020)	PRJNA418053		Reptiles	-	11	0
(Al-Masaudi et al., 2017)	PRJNA486341	28473812	Goat, Camel	-	12	0
-	PRJNA526439	-	Duck	-	8	0
-	PRJNA526291	-	Duck, Goose	-	8	0
-	PRJNA340484	-	Gorilla	-	1	0
-	PRJNA340483	-	Gorilla	-	1	0
-	PRJNA540280	-	Tortoise	-	1	1
-	PRJNA545614	-	Elephant	-	1	0
-	PRJEB23356	-	Chicken	-	651	55
(Cao et al., 2020)	PRJNA556790, PRJNA563508	32122398	Bird, human	CaoJ	158	17
-	PRJNA427653	-	Cow, Sheep, Goat, Horse, Human, Soil	-	81	24
(D'Arc et al., 2018)	PRJNA419744	29402305	Gorilla	DarcM	23	5
-	PRJNA579034	-	Sea cucumber	-	3	0
(Kamke et al., 2016)	PRJNA202380	27760570	Sheep	KamkeJ	20	20
(Lima et al., 2019)	PRJEB21624	31440274	Cattle	LimaJ	45	56
NeumannA	PRJNA340521, PRJNA340522, PRJNA340523, PRJNA340484, PRJNA340483, PRJNA340524, PRJNA340485, PRJNA340486, PRJNA340490, PRJNA340488, PRJNA340489, PRJNA340487	-	Capybara, lemur, gorilla, Orangutan, colobus	NeumannA	12	7
(Poulsen, 2019)	PRJEB31650	-	Pig		311	9
(Zaheer et al., 2019)	PRJNA529711, PRJNA420682	29651035	Bovine		41	28
(Ang et al., 2020)	PRJNA574841	32523128	Langur spp., Surili	AngA	11	2
(H. Wang et al., 2019)	PRJNA483083	30800107	Rhesus macaque	WangH	16	9
(Namasivayam et al., 2019)	PRJNA541010	31164469	Rhesus macaque	NamazivayamS	24	2
(Wallace et al., 2015)	PRJEB10338	26494241	Bovine		8	9
(Auffret et al., 2017)	PRJEB21624	29375511	Bovine		45	58
(Yang et al., 2016)	PRJNA292471	26873315	Bovine		24	9
(Huebner et al., 2019)	PRJNA453374	30796295	Bovine		26	8

-	PRJNA373898	-	Sheep		3	0
(Pollock et al., 2020)	PRJEB34736	32015392	Pig		35	22
-	PRJNA644746	-	Ruemn liquid	PRJNA644746	15	1
(Wolff et al., 2017)	PRJNA214227	28595639	Sheep, Red deer, cattle, whitetail, bison	WolffSM	48	27
(Xue et al., 2020)	PRJNA601318	32972462	Sheep	XueY	10	3
(Ye et al., 2018)	PRJNA356225	30077182	Human	YeZ	150	1
(Zhou et al., 2020)	PRJEB29373	31806420	Human	-	134	0
(Weng et al., 2019)	PRJNA429990	31240835	Human	-	40	0
(Guitor et al., 2019)	PRJNA540073	31611361	Human	GuitorA	6	1
(Peters et al., 2019)	PRJNA541981	31597568	Human	PetersB	59	1
(Philips et al., 2017)	PRJNA354503	28609785	Human	-	165	0
(Hall et al., 2017)	PRJNA385949	29183332	Human	-	262	0
-	PRJEB7949	-	Human	ERR69562(X)	40	4
(Wang et al., 2018)	ERP010708	30208875	Human	Wang	201	49
(Franzosa et al., 2019)	PRJNA400072	30531976	Human	Franzosa	270	20
(Araos et al., 2019)	PRJNA531921	31611867	Human	AraosR	79	26
(Lokmer et al., 2019)	PRJEB27005	30726303	Human	LokmerA	57	7
(Lloyd-Price et al., 2019)	PRJNA398089	31142855	Human	Price	1338	11
(Tett et al., 2019)	PRJNA529400	31607556	Human	TettAJ	162	9

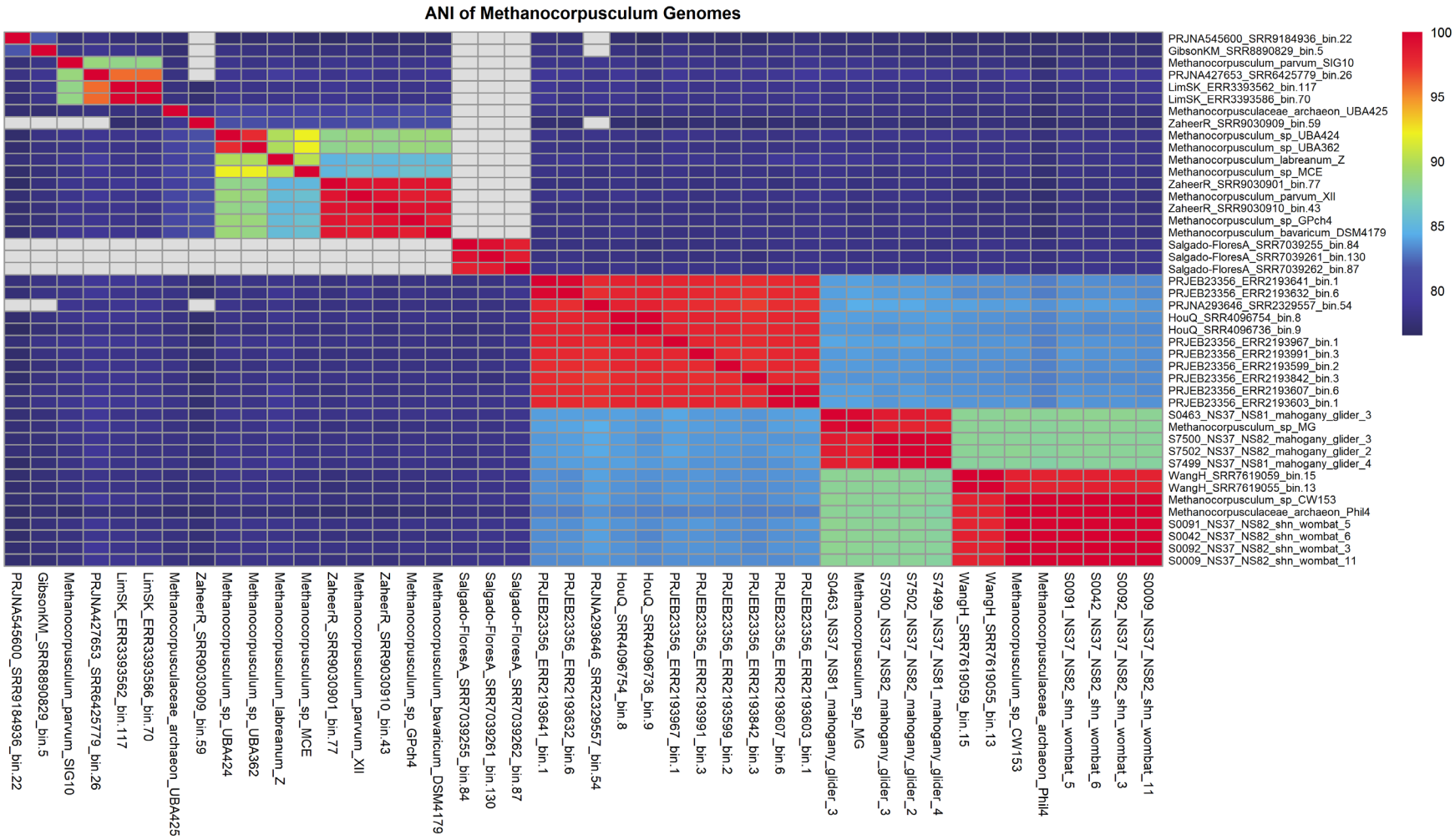


Figure 6.4. Average nucleotide identity (ANI) of high-quality (HQ) *Methanocorpusculum* MAGs and isolate genomes. ANI was calculated using FastANI (v1.1) and visualised using the pheatmaps and ggplot packages in R studio. ANI values are coloured according to the legend, where blank cells represent values with <70% identity.

Table 6.16. Basic genome statistics of *Methanocorpusculum* MAGs and isolate genomes included in this study. Estimated completeness, contamination, and genome size of each genome assembly was determined using CheckM (v1.1.2). GTDB-tk (v1.3.0) was used for taxonomic classification.

Bin Id	Completeness	Contamination	Strain heterogeneity	Coverage (x)	Genome size	Genome Type	GTDB classification	Source	Geography	Recovered in this study
CaoJ_2020_SRR10083254_bin.53	68.72	0	0	1.61	1759377	MAG	g__Methanocorpusculum;s__	Livestock	China	Y
CaoJ_2020_SRR10083263_bin.91	55.32	2.34	100	14.04	2478619	MAG	g__Methanocorpusculum;s__	Livestock	China	Y
CaoJ_2020_SRR10083264_bin.121	61.43	2.8	50	10.29	1602201	MAG	g__Methanocorpusculum;s__	Livestock	China	Y
DosterE_SRR3181107_bin.55	54.98	0	0	3.95	2271669	MAG	g__Methanocorpusculum;s__	Bos taurus	USA	Y
DosterE_SRR3181110_bin.36	74.03	0.98	100	4.19	1308083	MAG	g__Methanocorpusculum;s__	Bos taurus	USA	Y
DosterE_SRR3181120_bin.48	64.26	0.65	0	3.31	2148399	MAG	g__Methanocorpusculum;s__	Bos taurus	USA	Y
GibsonKM_SRR8890819_bin.52	60.2	3.1	0	4.50	1397535	MAG	g__Methanocorpusculum;s__	Rhino	South Africa	Y
GibsonKM_SRR8890819_bin.66	73.69	1.31	100	5.25	2174111	MAG	f__Methanocorpusculaceae;g__;s__	Rhino	South Africa	Y
GibsonKM_SRR8890819_bin.74	51.6	4.25	11.11	3.46	2302000	MAG	f__Methanocorpusculaceae;g__;s__	Rhino	South Africa	Y
GibsonKM_SRR8890821_bin.6	51.37	0	0	4.58	2607736	MAG	f__Methanocorpusculaceae;g__;s__	Rhino	South Africa	Y
GibsonKM_SRR8890827_bin.12	66.67	1.31	33.33	9.93	1666960	MAG	g__Methanocorpusculum;s__	Rhino	South Africa	Y
GibsonKM_SRR8890828_bin.11	60.29	0.65	0	13.33	1491482	MAG	f__Methanocorpusculaceae;g__;s__	Rhino	South Africa	Y
GibsonKM_SRR8890829_bin.5	92.55	0	0	7.85	1501125	MAG	g__Methanocorpusculum;s__	Rhino	South Africa	Y
GibsonKM_SRR8890833_bin.44	60.29	0.65	0	16.87	1926769	MAG	f__Methanocorpusculaceae;g__;s__	Rhino	South Africa	Y
GibsonKM_SRR8890836_bin.17	88.89	0.65	0	59.81	1957281	MAG	g__Methanocorpusculum;s__	Rhino	USA	Y
HouQ_SRR4096732_bin.27	82.82	2.61	25	6.53	2296936	MAG	g__Methanocorpusculum;s__	Chicken	China	Y
HouQ_SRR4096733_bin.42	74.13	1.63	0	6.08	1514216	MAG	g__Methanocorpusculum;s__	Chicken	China	Y
HouQ_SRR4096735_bin.14	85.62	1.36	0	8.37	1853659	MAG	g__Methanocorpusculum;s__	Chicken	China	Y
HouQ_SRR4096736_bin.9	90.52	1.31	0	23.34	2191231	MAG	g__Methanocorpusculum;s__	Chicken	China	Y
HouQ_SRR4096743_bin.14	80.8	3.16	0	6.91	2175834	MAG	g__Methanocorpusculum;s__	Chicken	China	Y
HouQ_SRR4096753_bin.75	78.89	0.65	0	5.95	1772587	MAG	g__Methanocorpusculum;s__	Chicken	China	Y
HouQ_SRR4096754_bin.8	96.41	1.31	0	13.76	2171828	MAG	g__Methanocorpusculum;s__	Chicken	China	Y
HouQ_SRR4096756_bin.20	76.18	1.31	0	8.54	1380501	MAG	g__Methanocorpusculum;s__	Chicken	China	Y
HouQ_SRR4096758_bin.27	71.9	0.65	0	10.30	1423311	MAG	g__Methanocorpusculum;s__	Chicken	China	Y
IlmbergerN_SRR1221442_bin.66	77.36	6.93	55.56	14.49	1365881	MAG	g__Methanocorpusculum;s__	Elephant	Germany	Y

LiC_2019_SRR6192929_bin.122	60.33	0.87	50	1317.52	1329138	MAG	g__Methanocorpusculum;s__	Sperm whale	China	Y
LimSK_ERR3393558_bin.66	66.97	1.96	66.67	5.11	792040	MAG	g__Methanocorpusculum;s__	Bos taurus	South Korea	Y
LimSK_ERR3393559_bin.18	80.39	0	0	5.32	1065924	MAG	g__Methanocorpusculum;s__	Bos taurus	South Korea	Y
LimSK_ERR3393562_bin.117	95.75	0.65	0	16.40	1472539	MAG	g__Methanocorpusculum;s__	Bos taurus	South Korea	Y
LimSK_ERR3393569_bin.13	67.55	6.59	78.57	11.04	988518	MAG	g__Methanocorpusculum;s__	Bos taurus	South Korea	Y
LimSK_ERR3393580_bin.39	56.03	0.65	100	5.38	661378	MAG	g__Methanocorpusculum;s__	Bos taurus	South Korea	Y
LimSK_ERR3393586_bin.70	95.75	0.65	0	12.43	1492000	MAG	g__Methanocorpusculum;s__	Bos taurus	South Korea	Y
Methanocorpusculaceae_archaeon_Phil4	91.85	2.29	40	184.00	1965334	MAG	s__Methanocorpusculum sp001940805	Wombat	Australia	N
Methanocorpusculaceae_archaeon_UBA425	97.71	0.65	0	477.30	1810882	MAG	s__Methanocorpusculum sp002506085	Mud fermentation metagenome	China	N
Methanocorpusculaceae_archaeon_UBA456	98.69	0	0	12.08	2074195	MAG	s__Methanocalculus sp002496395	Bioreactor metagenome	China	N
Methanocorpusculaceae_archaeon_WOFA02	99.35	0	0	105.27	1861543	MAG	s__Methanocalculus sp9912u	Oil field metagenome	Germany	N
Methanocorpusculum_bavaricum_DSM4179	98.21	0.66	0	-	1702624	Isolate	s__Methanocorpusculum parvum	Muddy sediment of wastewater pond	Germany	N
Methanocorpusculum_labreanum_Z	99.54	0	0	-	1804962	Isolate	s__Methanocorpusculum labreanum	Tar Pit Lake Sediment	USA	N
Methanocorpusculum_parvum_AS07pgkLD_39	96.7	7.64	64.71	301.95	1519932	MAG	s__Methanocorpusculum parvum	Anerobic digester	Denmark	N
Methanocorpusculum_parvum_SIG10	93.14	0.65	0	-	1423213	MAG	g__Methanocorpusculum;s__	Goat	USA	N
Methanocorpusculum_parvum_SIG11	86.88	1.31	50	-	1401745	MAG	g__Methanocorpusculum;s__	Goat	USA	N
Methanocorpusculum_parvum_XII	98.21	0	0	213.00	1709133	Isolate	s__Methanocorpusculum parvum	Anaerobic sour whey digester	Germany	N
Methanocorpusculum_sp_CW153	97.69	1.96	0	177.10	1944273	Isolate	s__Methanocorpusculum sp001940805	Wombat	Australia	Y
Methanocorpusculum_sp_GPch4	98.21	1.32	0	93.00	1915262	MAG	s__Methanocorpusculum parvum	Soil	Germany	N
Methanocorpusculum_sp_HGM00632	54.63	2.83	80	17.65	765679	MAG	g__Methanocorpusculum;s__	Human	Fiji	N
Methanocorpusculum_sp_MCE	93.58	0.66	100	900.00	1697696	MAG	s__Methanocorpusculum sp003315675	Metopus contortus-symbiont metagenome	Sweden	N
Methanocorpusculum_sp_MG	98.01	1.31	0	199.46	2026066	Isolate	g__Methanocorpusculum;s__	Mahogany glider	Australia	Y
Methanocorpusculum_sp_UBA362	96.86	0.66	100	37.29	1604060	MAG	s__Methanocorpusculum sp002498375	Mud fermentation metagenome	China	N
Methanocorpusculum_sp_UBA424	96.89	0	0	635.21	1561804	MAG	s__Methanocorpusculum sp002498375	Mud fermentation metagenome	China	N
Methanocorpusculum_sp_UBA592	86.37	0.66	0	19.01	1398626	MAG	s__Methanocorpusculum sp002498375	Mud fermentation metagenome	China	N
Nayfach_SRS475582_bin.12	81.64	4.84	88.89	17.70	1199510	MAG	g__Methanocorpusculum;s__	Human	Fiji	N
Nayfach_SRS476346_bin.34	83.91	2.61	75	10.71	1024803	MAG	g__Methanocorpusculum;s__	Human	Fiji	N
Nayfach_SRS476863_bin.43	74.67	2.97	66.67	9.80	986282	MAG	g__Methanocorpusculum;s__	Human	Fiji	N
PRJEB23356_ERR2193583_bin.2	62.54	1.31	0	1.64	1828800	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193592_bin.1	89.14	1.96	0	1.54	1093115	MAG	g__Methanocorpusculum;s__	Chicken	India	Y

PRJEB23356_ERR2193596_bin.1	72.1	1.36	0	1.27	973566	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193599_bin.2	90.52	1.31	0	2.34	1138870	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193600_bin.3	86.76	1.96	0	1.90	1382491	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193603_bin.1	91.05	1.85	0	2.77	2291275	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193605_bin.4	67.8	0.65	0	1.45	1874742	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193607_bin.6	91.07	1.31	0	2.57	1328394	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193608_bin.1	60.92	0.65	0	1.10	1311793	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193609_bin.1	79.88	1.31	0	1.79	1096261	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193618_bin.2	79.17	0.65	0	0.88	1318130	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193627_bin.3	84.64	3.27	0	3.36	1883413	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193631_bin.2	67.97	2.78	60	2.94	1665527	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193632_bin.6	90.96	1.53	0	1.32	1220113	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193633_bin.5	73.08	1.99	0	0.68	1670058	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193634_bin.1	59.39	1.96	0	1.06	1162356	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193635_bin.1	77.87	3.59	0	1.37	1538197	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193636_bin.7	63.41	0.98	0	0.99	1511797	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193637_bin.2	59.9	0.65	0	0.76	913350	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193641_bin.1	93.79	1.63	0	1.82	1306029	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193680_bin.6	54.47	1.31	0	1.29	1596425	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193685_bin.2	53.26	0.65	0	1.54	1238818	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193689_bin.3	69.61	0.65	0	2.50	1615385	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193796_bin.9	87.58	2.01	25	2.63	1426131	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193801_bin.5	51.27	0.65	0	1.78	1272521	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193803_bin.4	63.48	0.65	0	1.93	1163083	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193838_bin.4	64.78	0.65	0	1.94	1616174	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193840_bin.1	81.62	1.53	0	2.50	1486055	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193842_bin.3	90.39	1.63	33.33	2.88	1610041	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193845_bin.3	73.42	2.99	16.67	2.16	1999647	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193899_bin.7	57.97	0.78	0	2.04	955048	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193907_bin.1	53	1.31	0	2.48	1145976	MAG	g__Methanocorpusculum;s__	Chicken	India	Y

PRJEB23356_ERR2193914_bin.3	66.13	0.65	0	3.68	1365518	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193965_bin.14	78.05	1.96	0	1.99	1272959	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193966_bin.12	78.89	1.96	0	2.09	1371999	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193967_bin.1	90.5	1.31	0	4.52	1826713	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193968_bin.5	58.63	0.16	100	2.06	1904877	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193969_bin.4	81.67	2.09	0	2.51	1655682	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193971_bin.6	54.9	0.65	0	1.63	1918334	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193974_bin.4	52.87	0.65	0	1.21	892621	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193975_bin.2	54.32	0.65	0	2.03	959524	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193978_bin.3	66.1	0.65	0	2.48	1563639	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193986_bin.2	60.57	0.65	0	2.25	1780486	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193987_bin.6	86.6	1.96	0	4.46	1320431	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193991_bin.3	94.42	2.94	40	3.54	2617974	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193992_bin.6	88.7	1.44	0	2.19	1153668	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193993_bin.11	63.51	0.78	0	2.10	1062946	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193996_bin.1	86.89	1.63	0	3.44	1166539	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2193998_bin.4	86.9	1.96	33.33	2.43	1412596	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2194004_bin.4	50.25	0.05	0	1.62	1489449	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2194032_bin.3	76.37	3.54	42.86	2.77	1070736	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2194157_bin.9	79.39	0	0	1.89	928196	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2194196_bin.5	73.7	0.65	0	1.52	1340563	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2194200_bin.7	66.81	1.96	0	0.80	1460335	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJEB23356_ERR2194201_bin.5	57.4	2.75	20	0.81	1883708	MAG	g__Methanocorpusculum;s__	Chicken	India	Y
PRJNA293646_SRR2329557_bin.54	90.35	2.09	75	5.04	1396545	MAG	g__Methanocorpusculum;s__	Chicken	China	Y
PRJNA427653_SRR6425746_bin.31	65.01	7.84	93.33	15.59	953411	MAG	g__Methanocorpusculum;s__	Horse	China	Y
PRJNA427653_SRR6425771_bin.14	83.92	3.51	42.86	8.59	1332313	MAG	g__Methanocorpusculum;s__	Sheep	China	Y
PRJNA427653_SRR6425774_bin.39	74.62	6.09	68.75	11.19	1456948	MAG	g__Methanocorpusculum;s__	Sheep	China	Y
PRJNA427653_SRR6425779_bin.26	90.52	1.31	33.33	13.51	1473264	MAG	g__Methanocorpusculum;s__	Cow	China	Y
PRJNA427653_SRR6425780_bin.2	69.2	1.01	100	6.82	1096327	MAG	g__Methanocorpusculum;s__	Cow	China	Y
PRJNA427653_SRR6425794_bin.25	53.81	3.27	25	6.54	1256001	MAG	g__Methanocorpusculum;s__	Sheep	China	Y

PRJNA427653_SRR6425795_bin.15	67.49	1.86	33.33	6.52	1568506	MAG	g__Methanocorpusculum;s__	Sheep	China	Y
PRJNA427653_SRR6425800_bin.36	80.18	1.31	33.33	10.33	1025032	MAG	g__Methanocorpusculum;s__	Horse	China	Y
PRJNA427653_SRR6425803_bin.57	73.74	1.31	100	7.59	1287024	MAG	g__Methanocorpusculum;s__	Goat	China	Y
PRJNA427653_SRR6425807_bin.15	73.67	1.96	66.67	6.36	1335828	MAG	g__Methanocorpusculum;s__	Goat	China	Y
PRJNA427653_SRR6425821_bin.86	89.98	1.31	100	20.45	1080590	MAG	g__Methanocorpusculum;s__	Sheep	China	Y
PRJNA427653_SRR6425822_bin.5	84.63	1.85	80	8.38	1550856	MAG	g__Methanocorpusculum;s__	Sheep	China	Y
PRJNA427653_SRR6425824_bin.11	83.76	0.9	100	7.80	1048764	MAG	g__Methanocorpusculum;s__	Cow	China	Y
PRJNA545600_SRR9184936_bin.21	67.13	0.93	100	7.68	1567192	MAG	f__Methanocorpusculaceae;g__;s__	Elephant	China	Y
PRJNA545600_SRR9184936_bin.22	94.72	0	0	15.27	952996	MAG	g__Methanocorpusculum;s__	Elephant	China	Y
PRJNA545601_SRR9184932_bin.19	51.71	1.87	100	5.75	825446	MAG	f__Methanocorpusculaceae;g__;s__	Elephant	China	Y
PRJNA545606_SRR9184944_bin.3	87.85	2.61	25	5.73	1299820	MAG	g__Methanocorpusculum;s__	Elephant	China	Y
PRJNA545606_SRR9184944_bin.35	76.24	1.31	0	6.89	1400962	MAG	f__Methanocorpusculaceae;g__;s__	Elephant	China	Y
PRJNA590977_SRR10505693_bin.62	74.77	1.31	0	30.88	1157179	MAG	f__Methanocorpusculaceae;g__;s__	Horse	United Kingdom	Y
PRJNA590977_SRR10505694_bin.46	87.58	3.1	12.5	9.56	1231861	MAG	g__Methanocorpusculum;s__	Horse	United Kingdom	Y
PRJNA590977_SRR10505695_bin.11	80.1	3.49	30	34.82	1001694	MAG	f__Methanocorpusculaceae;g__;s__	Horse	United Kingdom	Y
PRJNA590977_SRR10505696_bin.31	80.79	4.58	14.29	16.82	1570625	MAG	f__Methanocorpusculaceae;g__;s__	Horse	United Kingdom	Y
PRJNA590977_SRR10505697_bin.33	86.9	3.43	33.33	12.44	1911395	MAG	f__Methanocorpusculaceae;g__;s__	Horse	United Kingdom	Y
S0009_NS37_NS82_shn_wombat_11	95.59	1.96	0	65.20	1858179	MAG	s__Methanocorpusculum sp001940805	Wombat	Australia	Y
S0041_NS37_NS82_shn_wombat_29	88.53	1.99	25	43.90	1878269	MAG	s__Methanocorpusculum sp001940805	Wombat	Australia	Y
S0041_NS37_NS82_shn_wombat_39	83.51	4.58	37.5	30.30	1962096	MAG	g__Methanocorpusculum;s__	Wombat	Australia	Y
S0042_NS37_NS82_shn_wombat_6	93.11	2.61	0	41.20	1802373	MAG	s__Methanocorpusculum sp001940805	Wombat	Australia	Y
S0091_NS37_NS82_shn_wombat_5	95.1	1.96	0	12.40	1830852	MAG	s__Methanocorpusculum sp001940805	Wombat	Australia	Y
S0092_NS37_NS82_shn_wombat_3	96.38	1.96	0	25.20	1904683	MAG	s__Methanocorpusculum sp001940805	Wombat	Australia	Y
S0463_NS37_NS81_mahogany_glider_3	94.42	1.96	50	16.50	1706078	MAG	g__Methanocorpusculum;s__	Mahogany glider	Australia	Y
S7499_NS37_NS81_mahogany_glider_4	90.17	1.96	0	10.00	1694480	MAG	g__Methanocorpusculum;s__	Mahogany glider	Australia	Y
S7500_NS37_NS82_mahogany_glider_3	97.39	1.31	0	22.20	1959252	MAG	g__Methanocorpusculum;s__	Mahogany glider	Australia	Y
S7502_NS37_NS82_mahogany_glider_2	96.41	1.31	0	20.30	1937944	MAG	g__Methanocorpusculum;s__	Mahogany glider	Australia	Y
Salgado-FloresA_SRR7039255_bin.84	91.18	1.96	0	6.80	886206	MAG	g__Methanocorpusculum;s__	Ptarmigan	Norway	Y
Salgado-FloresA_SRR7039256_bins.9	66.24	1.96	33.33	4.77	1012566	MAG	g__Methanocorpusculum;s__	Ptarmigan	Norway	Y
Salgado-FloresA_SRR7039257_bins.17	77.76	1.55	50	5.31	929179	MAG	g__Methanocorpusculum;s__	Ptarmigan	Norway	Y

Salgado-FloresA_SRR7039258_bin.77	89.87	0.65	0	13.17	1322197	MAG	g__Methanocorpusculum;s__	Ptarmigan	Norway	Y
Salgado-FloresA_SRR7039261_bin.130	91.83	0.65	0	9.36	1109334	MAG	g__Methanocorpusculum;s__	Ptarmigan	Norway	Y
Salgado-FloresA_SRR7039262_bin.87	90.52	1.36	33.33	9.90	1563200	MAG	g__Methanocorpusculum;s__	Ptarmigan	Norway	Y
SanzPR_SRR5678935_bin.21	56.22	0	0	3.28	723304	MAG	g__Methanocorpusculum;s__	Bos taurus	USA	Y
SanzPR_SRR5678937_bin.46	51.98	2.29	0	2.76	980181	MAG	g__Methanocorpusculum;s__	Bos taurus	USA	Y
SanzPR_SRR5678984_bin.64	82.1	2.83	66.67	8.07	1299004	MAG	g__Methanocorpusculum;s__	Bos taurus	Canada	Y
SanzPR_SRR5679020_bin.72	83.05	1.96	33.33	3.73	986764	MAG	g__Methanocorpusculum;s__	Bos taurus	Canada	Y
WangH_SRR7619055_bin.13	97.69	1.96	0	35.16	829087	MAG	s__Methanocorpusculum sp001940805	Rhesus monkey	China	Y
WangH_SRR7619059_bin.15	94.42	1.96	0	18.26	851558	MAG	s__Methanocorpusculum sp001940805	Rhesus monkey	China	Y
ZaheerR_SRR9030901_bin.77	90.09	2.98	80	13.68	665848	MAG	s__Methanocorpusculum parvum	feedlot catch-basin water	Canada	Y
ZaheerR_SRR9030902_bin.4	77.99	1.96	66.67	51.35	717580	MAG	g__Methanocorpusculum;s__	feedlot catch-basin water	Canada	Y
ZaheerR_SRR9030902_bin.64	88.77	1.96	75	30.15	628379	MAG	g__Methanocorpusculum;s__	feedlot catch-basin water	Canada	Y
ZaheerR_SRR9030909_bin.59	90.98	3.32	90	17.46	1515459	MAG	g__Methanocorpusculum;s__	feedlot catch-basin water	Canada	Y
ZaheerR_SRR9030910_bin.43	90.41	2.32	80	19.76	677014	MAG	s__Methanocorpusculum parvum	feedlot catch-basin water	Canada	Y
ZaheerR_SRR9030910_bin.98	88.97	7.54	57.14	13.53	707438	MAG	g__Methanocorpusculum;s__	feedlot catch-basin water	Canada	Y
ZaheerR_SRR9037495_bin.33	83.28	2.29	75	7.20	1249300	MAG	g__Methanocorpusculum;s__	Bos taurus	Canada	Y

Table 6.17. Core, accessory, unique and exclusively absent genes of high-quality (HQ) *Methanocorpusculum* genomes. The gene counts were determined using BPGA (Chaudhari et al., 2016). Gene families were clustered using USEARCH, with the default 50% sequence identity cut-off for orthologous clustering.

Organism name	Core genes	Accessory genes	Unique genes	Exclusively absent genes
GibsonKM_SRR8890829_bin.5	180	1085	260	3
HouQ_SRR4096736_bin.9	180	1173	0	22
HouQ_SRR4096754_bin.8	180	1357	22	0
LimSK_ERR3393562_bin.117	180	1284	25	0
LimSK_ERR3393586_bin.70	180	1294	38	0
Methanocorpusculaceae_archaeon_Phil4	180	1579	143	14
Methanocorpusculaceae_archaeon_UBA425	180	1354	196	1
Methanocorpusculum_bavarium_DSM4179	180	1433	54	0
Methanocorpusculum_labreanum_Z	180	1460	92	1
Methanocorpusculum_parvum_SIG10	180	1162	114	5
Methanocorpusculum_parvum_XII	180	1437	49	1
Methanocorpusculum_sp_CW153	180	1672	8	0
Methanocorpusculum_sp_GPch4	180	1443	195	0
Methanocorpusculum_sp_MCE	180	1321	117	2
Methanocorpusculum_sp_MG	180	1712	98	0
Methanocorpusculum_sp_UBA362	180	1333	96	4
Methanocorpusculum_sp_UBA424	180	1316	48	2
PRJEB23356_ERR2193599_bin.2	180	1290	51	9
PRJEB23356_ERR2193603_bin.1	180	1284	45	5
PRJEB23356_ERR2193607_bin.6	180	1233	41	10
PRJEB23356_ERR2193632_bin.6	180	1195	44	10
PRJEB23356_ERR2193641_bin.1	180	1234	22	4
PRJEB23356_ERR2193842_bin.3	180	1167	19	7
PRJEB23356_ERR2193967_bin.1	180	1261	20	1
PRJEB23356_ERR2193991_bin.3	180	1234	20	0
PRJNA293646_SRR2329557_bin.54	180	1039	66	28
PRJNA427653_SRR6425779_bin.26	180	1077	46	3
PRJNA545600_SRR9184936_bin.22	180	1111	289	2
S0009_NS37_NS82_shn_wombat_11	180	1623	8	3
S0042_NS37_NS82_shn_wombat_6	180	1578	3	0
S0091_NS37_NS82_shn_wombat_5	180	1598	37	0
S0092_NS37_NS82_shn_wombat_3	180	1639	9	0
S0463_NS37_NS81_mahogany_glider_3	180	1480	57	4
S7499_NS37_NS81_mahogany_glider_4	180	1513	68	12
S7500_NS37_NS82_mahogany_glider_3	180	1764	41	0
S7502_NS37_NS82_mahogany_glider_2	180	1733	25	0
Salgado-FloresA_SRR7039255_bin.84	180	726	28	9
Salgado-FloresA_SRR7039261_bin.130	180	755	14	4
Salgado-FloresA_SRR7039262_bin.87	180	727	41	7
WangH_SRR7619055_bin.13	180	1711	3	0
WangH_SRR7619059_bin.15	180	1701	8	0
ZaheerR_SRR9030901_bin.77	180	1230	61	4
ZaheerR_SRR9030909_bin.59	180	1149	103	11
ZaheerR_SRR9030910_bin.43	180	1209	53	4

Table 6.18. KEGG Orthologs (KO) enriched in environmental (Env) and host-associated (HA) *Methanocorpusculum* genomes. KO annotation and statistical analysis was performed using 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). Only KOs with corrected p values of <0.05 were retained and considered significant, as determined by Fisher's Exact Test. 33 HA and 11 Env genomes were included in the analysis.

KO	Env Gen.	Host Gen.	Corrected P Value	Name	Description
K09011	7	2	0.000480882	cimA	(R)-citramalate synthase [EC:2.3.1.182]
K03790	5	1	0.006512307	rimJ	[ribosomal protein S5]-alanine N-acetyltransferase [EC:2.3.1.267]
K19002	3	0	0.048001611	mgs, bgsB	1,2-diacylglycerol 3-alpha-glucosyltransferase [EC:2.4.1.337]
K13678	3	0	0.048001611	cpoA	1,2-diacylglycerol-3-alpha-glucose alpha-1,2-galactosyltransferase [EC:2.4.1.-]
K13677	5	2	0.014662646	dgs, bgsA	1,2-diacylglycerol-3-alpha-glucose alpha-1,2-glucosyltransferase [EC:2.4.1.208]
K14652	8	7	0.001396109	ribBA	3,4-dihydroxy 2-butanone 4-phosphate synthase / GTP cyclohydrolase II [EC:4.1.99.12 3.5.4.25]
K02858	0	24	0.011605847	ribB, RIB3	3,4-dihydroxy 2-butanone 4-phosphate synthase [EC:4.1.99.12]
K01607	0	24	0.011605847	pcaC	4-carboxymuconolactone decarboxylase [EC:4.1.1.44]
K01666	4	0	0.010076308	mhpE	4-hydroxy 2-oxovalerate aldolase [EC:4.1.3.39]
K10713	0	22	0.046250597	fae	5,6,7,8-tetrahydromethanopterin hydro-lyase [EC:4.2.1.147]
K01066	0	24	0.011605847	aes	acetyl esterase [EC:3.1.1.-]
K00197	6	2	0.003173256	cdhE, acsC	acetyl-CoA decarbonylase/synthase complex subunit gamma [EC:2.1.1.245]
K07107	0	24	0.011605847	ybgC	acyl-CoA thioester hydrolase [EC:3.1.2.-]
K00759	8	3	0.000184955	APRT, apt	adenine phosphoribosyltransferase [EC:2.4.2.7]
K02232	0	25	0.009077431	cobQ, cbiP	adenosylcobyric acid synthase [EC:6.3.5.10]
K00002	7	0	7.57E-05	AKR1A1, adh	alcohol dehydrogenase (NADP+) [EC:1.1.1.2]
K03738	5	1	0.006512307	aor	aldehyde:ferredoxin oxidoreductase [EC:1.2.7.5]
K18815	4	0	0.010076308	aac6-I	aminoglycoside 6'-N-acetyltransferase I [EC:2.3.1.82]
K00385	6	1	0.001215527	asrC	anaerobic sulfite reductase subunit C
K00766	4	0	0.010076308	trpD	anthranilate phosphoribosyltransferase [EC:2.4.2.18]
K01657	3	0	0.048001611	trpE	anthranilate synthase component I [EC:4.1.3.27]
K01658	4	0	0.010076308	trpG	anthranilate synthase component II [EC:4.1.3.27]
K04749	6	3	0.006512307	rsbV	anti-sigma B factor antagonist

K01584	0	22	0.046250597	adiA	arginine decarboxylase [EC:4.1.1.19]
K03325	7	2	0.000480882	ACR3, arsB	arsenite transporter
K06989	6	0	0.000480882	nadX, ASPDH	aspartate dehydrogenase [EC:1.4.1.21]
K06984	0	22	0.046250597	K06984	beta-ribofuranosylaminobenzene 5'-phosphate synthase [EC:2.4.2.54]
K04042	8	10	0.006131247	glmU	bifunctional UDP-N-acetylglucosamine pyrophosphorylase [EC:2.7.7.23 2.3.1.157]
K01537	7	2	0.000480882	E3.6.3.8	Ca2+-transporting ATPase [EC:3.6.3.8]
K14170	0	28	0.002473108	pheA	chorismate mutase / prephenate dehydratase [EC:5.4.99.5 4.2.1.51]
K08482	6	2	0.003173256	kaiC	circadian clock protein KaiC
K18431	3	0	0.048001611	legF, ptmB	CMP-N,N'-diacetyllegionaminic acid synthase [EC:2.7.7.82]
K19817	3	33	0.032129681	fprA	coenzyme F420H2 oxidase [EC:1.5.3.22]
K07012	8	14	0.046250597	cas3	CRISPR-associated endonuclease/helicase Cas3 [EC:3.1.-.- 3.6.4.-]
K09951	8	13	0.016292824	cas2	CRISPR-associated protein Cas2
K07725	0	25	0.009077431	csa3	CRISPR-associated protein Csa3
K01533	0	24	0.011605847	copB	Cu2+-exporting ATPase [EC:3.6.3.4]
K03146	7	2	0.000480882	THI4, THI1	cysteine-dependent adenosine diphosphate thiazole synthase [EC:2.4.2.60]
K01485	8	2	7.57E-05	codA	cytosine/creatinine deaminase [EC:3.5.4.1 3.5.4.21]
K00657	0	29	0.001396109	speG, SAT	diamine N-acetyltransferase [EC:2.3.1.57]
K01270	0	30	0.000918048	pepD	dipeptidase D [EC:3.4.13.-]
K07343	8	7	0.001396109	tfoX	DNA transformation protein and related proteins
K01247	0	27	0.0037779	alkA	DNA-3-methyladenine glycosylase II [EC:3.2.2.21]
K03298	8	3	0.000184955	TC.DME	drug/metabolite transporter, DME family
K02217	0	32	0.000416148	ftnA, ftn	ferritin [EC:1.16.3.2]
K18429	3	0	0.048001611	legG, neuC2	GDP/UDP-N,N'-diacetylbacillosamine 2-epimerase (hydrolysing) [EC:3.2.1.184]
K00046	7	4	0.002253084	idnO	gluconate 5-dehydrogenase [EC:1.1.1.69]
K00820	8	9	0.0037779	glmS, GFPT	glucosamine---fructose-6-phosphate aminotransferase (isomerizing) [EC:2.6.1.16]
K00973	8	13	0.016292824	rfbA, rffH	glucose-1-phosphate thymidylyltransferase [EC:2.7.7.24]
K02000	7	3	0.001109728	proV	glycine betaine/proline transport system ATP-binding protein [EC:3.6.3.32]
K02001	7	3	0.001109728	proW	glycine betaine/proline transport system permease protein

K02002	6	3	0.006512307	proX	glycine betaine/proline transport system substrate-binding protein
K00641	8	13	0.016292824	metX	homoserine O-acetyltransferase/O-succinyltransferase [EC:2.3.1.31 2.3.1.46]
K01749	1	31	0.0037779	hemC, HMBS	hydroxymethylbilane synthase [EC:2.5.1.61]
K00760	0	32	0.000416148	hprT, hpt, HPRT1	hypoxanthine phosphoribosyltransferase [EC:2.4.2.8]
K01609	3	0	0.048001611	trpC	indole-3-glycerol phosphate synthase [EC:4.1.1.48]
K01843	8	2	7.57E-05	kamA	lysine 2,3-aminomutase [EC:5.4.3.2]
K16792	8	5	0.000480882	aksD	methanogen homoaconitase large subunit [EC:4.2.1.114]
K16793	8	5	0.000480882	aksE	methanogen homoaconitase small subunit [EC:4.2.1.114]
K10977	8	9	0.0037779	aksA	methanogen homocitrate synthase [EC:2.3.3.14 2.3.3.-]
K10978	8	5	0.000480882	aksF	methanogen homoisocitrate dehydrogenase [EC:1.1.1.87 1.1.1.-]
K08168	1	27	0.0260571	tetB	MFS transporter, DHA2 family, metal-tetracycline-proton antiporter
K18326	4	1	0.032129681	mdtD	MFS transporter, DHA2 family, multidrug resistance protein
K15497	5	2	0.014662646	wtpC	molybdate/tungstate transport system ATP-binding protein [EC:3.6.3.-3.6.3.55]
K15496	5	2	0.014662646	wtpB	molybdate/tungstate transport system permease protein
K15495	5	2	0.014662646	wtpA	molybdate/tungstate transport system substrate-binding protein
K03924	7	5	0.0037779	moxR	MoxR-like ATPase [EC:3.6.3.-]
K16052	0	27	0.0037779	ynaI, mscMJ	MscS family membrane protein
K18430	3	0	0.048001611	legI, neuB2	N,N'-diacetyllegionaminate synthase [EC:2.5.1.101]
K03308	0	30	0.000918048	TC.NSS	neurotransmitter:Na+ symporter, NSS family
K02051	7	5	0.0037779	ABC.SN.S	NitT/TauT family transport system substrate-binding protein
K01740	8	13	0.016292824	metY	O-acetylhomoserine (thiol)-lyase [EC:2.5.1.49]
K10947	0	21	0.048001611	padR	PadR family transcriptional regulator, regulatory protein PadR
K12267	8	15	0.048001611	msrAB	peptide methionine sulfoxide reductase msrA/msrB [EC:1.8.4.11 1.8.4.12]
K02032	8	10	0.006131247	ABC.PE.A1	peptide/nickel transport system ATP-binding protein
K02031	7	10	0.041517957	ABC.PE.A	peptide/nickel transport system ATP-binding protein
K02617	8	6	0.000918048	paaY	phenylacetic acid degradation protein
K03431	8	9	0.0037779	glmM	phosphoglucosamine mutase [EC:5.4.2.10]
K01091	8	2	7.57E-05	gph	phosphoglycolate phosphatase [EC:3.1.3.18]
K11725	0	23	0.016292824	LHPP	phospholysine phosphohistidine inorganic pyrophosphate phosphatase [EC:3.6.1.1 3.1.3.-]

K07315	7	2	0.000480882	rsbU_P	phosphoserine phosphatase RsbU/P [EC:3.1.3.3]
K04518	7	3	0.001109728	pheA2	prephenate dehydratase [EC:4.2.1.51]
K04517	6	2	0.003173256	tyrA2	prephenate dehydrogenase [EC:1.3.1.12]
K06959	0	24	0.011605847	tex	protein Tex
K06901	0	23	0.016292824	pbuG	putative MFS transporter, AGZA family, xanthine/uracil permease
K15045	3	0	0.048001611	RSAD2	radical S-adenosyl methionine domain-containing protein 2
K03538	8	15	0.048001611	POP4, RPP29	ribonuclease P protein subunit POP4 [EC:3.1.26.5]
K06384	0	24	0.011605847	spoIIM	stage II sporulation protein M
K10761	7	9	0.0260571	THG1	tRNA(His) guanylyltransferase [EC:2.7.7.79]
K01695	4	0	0.010076308	trpA	tryptophan synthase alpha chain [EC:4.2.1.20]
K01696	3	0	0.048001611	trpB	tryptophan synthase beta chain [EC:4.2.1.20]
K06857	8	13	0.016292824	tupC, vupC	tungstate transport system ATP-binding protein [EC:3.6.3.55]
K05773	8	12	0.011605847	tupB, vupB	tungstate transport system permease protein
K05772	8	12	0.011605847	tupA, vupA	tungstate transport system substrate-binding protein
K01791	5	1	0.006512307	wecB	UDP-N-acetylglucosamine 2-epimerase (non-hydrolysing) [EC:5.1.3.14]
K06889	8	5	0.000480882	K06889	uncharacterized protein
K07079	0	30	0.000918048	K07079	uncharacterized protein
K07112	0	29	0.001396109	K07112	uncharacterized protein
K06911	7	4	0.002253084	K06911	uncharacterized protein
K07003	1	29	0.011494634	K07003	uncharacterized protein
K09135	0	24	0.011605847	K09135	uncharacterized protein
K09125	0	24	0.011605847	K09125	uncharacterized protein
K02824	0	29	0.001396109	pyrP, uraA	uracil permease
K09817	0	28	0.002473108	znuC	zinc transport system ATP-binding protein [EC:3.6.3.-]
K09816	0	25	0.009077431	znuB	zinc transport system permease protein

Table 6.19. KEGG Orthologs (KO) enriched in environmental (Env) and host-associated (HA) *Methanocorpusculum* genomes. KO annotation and statistical analysis was performed using 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). Only KOs with corrected p values of <0.05 were retained and considered significant, as determined by Mann-Whitney U Test. 33 HA and 11 Env genomes were included in the analysis.

KO	Env Gene Count	Host Gene Count	P value	Corrected P Value	Gene
K01091	1.125	0.055555556	5.86E-09	4.12E-06	gph
K01843	1.625	0.083333333	6.91E-09	4.12E-06	kamA
K03325	1	0.055555556	1.33E-07	3.10E-05	ACR3, arsB
K07315	0.875	0.055555556	1.56E-07	3.10E-05	rsbU_P
K09011	0.875	0.055555556	1.56E-07	3.10E-05	cimA
K01537	1	0.055555556	1.33E-07	3.10E-05	E3.6.3.8
K00385	0.75	0.027777778	3.35E-07	4.89E-05	asrC
K00952	1.75	0.944444444	3.39E-07	4.89E-05	nadM
K03146	0.875	0.083333333	3.69E-07	4.89E-05	THI4, THI1
K06988	2.125	0.916666667	9.01E-07	7.32E-05	fno
K02000	0.875	0.083333333	9.83E-07	7.32E-05	proV
K01698	2.25	1	8.47E-07	7.32E-05	hemB, ALAD
K06927	1.875	1.083333333	9.83E-07	7.32E-05	DPH6
K02001	0.875	0.083333333	9.83E-07	7.32E-05	proW
K06889	1.125	0.138888889	6.85E-07	7.32E-05	K06889
K04518	0.875	0.083333333	9.83E-07	7.32E-05	pheA2
K08482	0.75	0.055555556	2.94E-06	0.000184142	kaiC
K00197	0.75	0.055555556	2.94E-06	0.000184142	cdhE, acsC
K04517	0.75	0.055555556	2.94E-06	0.000184142	tyrA2
K06911	0.875	0.111111111	4.47E-06	0.000266226	K06911
K19817	0.375	1.722222222	5.72E-06	0.000296228	fprA
K00046	2	0.25	5.67E-06	0.000296228	idnO
K03790	0.75	0.027777778	5.54E-06	0.000296228	rimJ
K03738	0.625	0.027777778	6.05E-06	0.000300326	aor
K01791	1.125	0.055555556	7.21E-06	0.000343796	wecB
K03924	1.75	0.25	1.35E-05	0.000619891	moxR
K02051	0.875	0.138888889	1.59E-05	0.000658932	ABC.SN.S
K01749	0.125	0.861111111	1.59E-05	0.000658932	hemC, HMBS
K02002	0.75	0.083333333	1.60E-05	0.000658932	proX
K09121	1.75	1.055555556	2.76E-05	0.001061888	larC
K00721	1.75	1.055555556	2.76E-05	0.001061888	DPM1
K04749	0.75	0.111111111	2.94E-05	0.001096644	rsbV
K02032	1.25	0.277777778	3.18E-05	0.001148898	ABC.PE.A1
K15497	0.625	0.055555556	4.55E-05	0.001464754	wtpC
K15496	0.625	0.055555556	4.55E-05	0.001464754	wtpB
K13677	0.625	0.055555556	4.55E-05	0.001464754	dgs, bgsA
K15495	0.625	0.055555556	4.55E-05	0.001464754	wtpA
K21029	1.75	1.027777778	6.94E-05	0.002175475	moeB

K03636	1.625	1.083333333	7.64E-05	0.002334166	moaD, cysO
K18326	0.5	0.027777778	9.37E-05	0.00279086	mdtD
K07003	0.125	1.444444444	0.000126238	0.003670137	K07003
K02379	1.875	2.666666667	0.00021452	0.006088288	fdhD
K16306	1.75	0.916666667	0.000228016	0.006320829	K16306
K01784	2.125	0.916666667	0.000270008	0.00731477	galE, GALE
K09951	1.25	0.388888889	0.000287354	0.007611682	cas2
K08963	1.625	1	0.000485329	0.012576352	mtnA
K10761	0.875	0.25	0.000541984	0.013745649	THG1
K07114	0.5	0.055555556	0.000577163	0.014332879	yfbK
K02031	1	0.277777778	0.000592039	0.014402244	ABC.PE.A
K02006	2.125	1.333333333	0.000641762	0.014433589	cbiO
K08168	0.125	0.777777778	0.000630363	0.014433589	tetB
K07089	1.625	0.75	0.000629427	0.014433589	K07089
K01817	1.375	0.944444444	0.000614999	0.014433589	trpF
K11717	1.375	2.472222222	0.00074954	0.016318617	sufS
K16785	2.5	1.527777778	0.000752956	0.016318617	ecfT
K12267	1.125	0.416666667	0.000782188	0.01664943	msrAB
K00973	1.25	0.444444444	0.000870376	0.018201556	E2.7.7.24, rfbA, rffH
K07130	1.375	0.916666667	0.000924058	0.01899098	kynB
K03655	1.75	0.416666667	0.000977951	0.019757925	recG
K07481	1	0.027777778	0.001135982	0.022198197	K07481
K19068	0.5	0.027777778	0.001135982	0.022198197	wbjC
K07504	0.375	0.027777778	0.001246381	0.023582321	K07504
K17716	0.375	0.027777778	0.001246381	0.023582321	capD
K07012	1.125	0.444444444	0.001574421	0.029323598	cas3
K01802	2.75	3.472222222	0.001622118	0.029747158	E5.2.1.8
K02013	4.375	3.166666667	0.001800179	0.032512323	ABC.FEV.A
K06223	0.125	0.722222222	0.001962664	0.03487054	dam
K07464	2	1.25	0.001989259	0.03487054	cas4
K19418	0.5	0.083333333	0.002137229	0.035881371	epsK
K00947	0.5	0.083333333	0.002137229	0.035881371	mosAB
K01126	0.5	0.083333333	0.002137229	0.035881371	E3.1.4.46, glpQ, ugpQ
K03458	0.25	0.777777778	0.00220654	0.036530496	TC.NCS2
K01501	0.875	1.527777778	0.002753356	0.044958908	E3.5.5.1
K09138	1.125	0.5	0.002874868	0.046308683	K09138

Table 6.20. KEGG Orthologs (KO) enriched in marsupial and non-marsupial host-associated (HA) *Methanocorpusculum* genomes. KO annotation and statistical analysis were performed using 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). Only KOs with corrected p values of <0.05 were retained and considered significant, as determined by Fisher's Exact Test. 11 marsupial and 22 non-marsupial HA genomes were included in the analysis.

KO	Marsupial Gen.	Host Gen.	Corrected P Value	Name	Description
K19055	10	2	4.47799E-05	prdX, proX	Ala-tRNA(Pro)deacylase [EC3.1.1.-]
K19954	5	0	0.024862614	adh1	alcohol dehydrogenase [EC1.1.1.-]
K19294	8	4	0.024122117	algI	alginate O-acetyltransferase complex protein AlgI
K19710	10	5	0.001428886	E2.7.7.53	ATP adenylyltransferase [EC2.7.7.53]
K16898	10	7	0.005398013	addA	ATP-dependent helicase/nuclease subunitA [EC3.1.-.-3.6.4.12]
K04114	9	3	0.002086743	bcrA, badF	benzoyl-CoA reductase subunitA [EC1.3.7.8]
K01012	10	2	4.47799E-05	bioB	biotinsynthase[EC2.8.1.6]
K01999	10	9	0.020309128	livK	branched-chain amino acid transport system substrate-binding protein
K03781	9	5	0.010513425	katE, CAT, catB, srpA	catalase [EC1.11.1.6]
K14136	8	4	0.024122117	K14136	decaprenyl-phosphate phosphoribosyl transferase [EC2.4.2.45]
K19548	7	2	0.020309128	bacC	dihydroanticapsinde hydrogenase[EC1.1.1.385]
K03658	8	2	0.004198421	helD	DNA helicaseIV [EC3.6.4.12]
K02348	0	15	0.048826882	elaA	ElaA protein
K11041	10	6	0.002573666	eta	exfoliative toxinA/B
K04758	0	17	0.014844536	feoA	ferrous iron transport protein A
K10563	7	2	0.020309128	mut	fpg; formamidopyrimidine-DNA glycosylase [EC3.2.2.234.2.99.18]
K00131	10	2	4.47799E-05	gapN	glyceraldehyde-3-phosphate dehydrogenase(NADP+)[EC1.2.1.9]
K00641	10	2	4.47799E-05	metX	homoserine O-acetyltransferase/O-succinyltransferase [EC2.3.1.312.3.1.46]
K14475	10	9	0.020309128	ICP	inhibitor of cysteine peptidase
K07271	10	1	4.47799E-05	licD	lipopolysaccharide choline phosphotransferase [EC2.7.8.-]

K18009	6	0	0.006305735	budC	meso-butanediol dehydrogenase [EC1.1.1.-1.1.1.76
K01740	10	2	4.47799E-05	metY	O-acetyl homoserine(thiol)-lyase [EC2.5.1.49]
K12267	9	5	0.010513425	msrAB	peptide methionine sulfoxide reductase msrA/msrB[EC1.8.4.111.8.4.12]
K02033	10	6	0.002573666	ABC.PE.P	peptide/nickel transport system permease protein
K02034	10	7	0.005398013	ABC.PE.P1	peptide/nickel transport system permease protein
K02035	10	6	0.002573666	ABC.PE.S	peptide/nickel transport system substrate-binding protein
K01841	7	0	0.001428886	pepM	phosphoenolpyruvate phosphomutase[EC5.4.2.9]
K09459	7	0	0.001428886	E4.1.1.82	phosphonopyruvate decarboxylase [EC4.1.1.82]
K03826	10	5	0.001428886	yiaC	putative acetyltransferase [EC2.3.1.-]
K07386	10	3	0.000174642	pepO	putative endopeptidase [EC3.4.24.-]
K03546	9	6	0.020309128	sbc	rad50; DNA repair protein SbcC/Rad50
K07574	10	2	4.47799E-05	yhbY	RNA-binding protein
K16868	5	0	0.024862614	tehB	Tellurite methyltransferase[EC2.1.1.265]
K07078	10	2	4.47799E-05	K07078	uncharacterized protein
K09790	10	2	4.47799E-05	K09790	uncharacterized protein
K09138	10	6	0.002573666	K09138	uncharacterized protein

Table 6.21. KEGG Orthologs (KO) enriched in marsupial and non-marsupial host-associated (HA) *Methanocorpusculum* genomes. KO annotation and statistical analysis were performed using 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). Only KOs with corrected p values of <0.05 were retained and considered significant, as determined by Mann-Whitney U Test. 11 marsupial and 22 non-marsupial HA genomes were included in the analysis.

KO	Marsupial Gene Count	Host Gene Count	pvalue	corrected_pvalue	Gene
K07271	1.7	0.04	1.98E-08	2.36E-05	licD
K07386	1.6	0.2	1.98E-06	0.000785865	pepO
K20265	1.9	0.44	1.80E-06	0.000785865	gadC
K08169	3.4	1.36	4.47E-06	0.001331809	yebQ
K04114	0.9	0.12	8.30E-06	0.001977718	bcrA, badF
K03658	0.8	0.08	1.49E-05	0.002960495	helD
K02003	2.8	1.2	1.93E-05	0.003289423	ABC.CD.A
K14475	1.6	0.4	3.12E-05	0.004642473	ICP
K00567	2	0.92	3.57E-05	0.004731758	ogt, MGMT
K07332	1.9	1	6.97E-05	0.007324666	flaI-A, flaI
K02016	6.9	3.68	6.15E-05	0.007324666	ABC.FEV.S
K03750	1.6	0.64	7.37E-05	0.007324666	moeA
K19548	0.7	0.08	0.000102556	0.008149809	bacC
K10563	0.7	0.08	0.000102556	0.008149809	mutM, fpg
K12267	0.9	0.2	9.12E-05	0.008149809	msrAB
K00059	2.2	0.8	0.000119086	0.008871933	fabG
K07333	1.9	1.04	0.000157417	0.011037698	flaJ-A, flaJ
K02004	2.8	1.52	0.000204221	0.013523951	ABC.CD.P
K03781	0.9	0.24	0.00021756	0.013649053	katE, CAT, catB, srpA
K03546	0.9	0.24	0.000239865	0.014295939	sbcC, rad50
K00200	3.2	1.84	0.000276224	0.01567898	fwdA, fmdA
K00202	3.2	1.88	0.000307263	0.016648092	fwdC, fmdC
K07124	0.6	0.08	0.000604131	0.020574986	K07124
K01546	0.6	0.08	0.000604131	0.020574986	kdpA
K02075	0.6	0.08	0.000604131	0.020574986	ABC.ZM.P
K07646	0.6	0.08	0.000604131	0.020574986	kdpD
K02435	0.4	0.92	0.000604131	0.020574986	gatC, GATC
K01547	0.6	0.08	0.000604131	0.020574986	kdpB
K00201	3.1	1.68	0.000546241	0.020574986	fwdB, fmdB
K03183	0.6	0.08	0.000604131	0.020574986	ubiE
K14136	0.8	0.2	0.000435421	0.020574986	K14136
K02077	0.6	0.08	0.000604131	0.020574986	ABC.ZM.S
K02074	0.6	0.08	0.000604131	0.020574986	ABC.ZM.A
K07133	3.4	2.4	0.000510647	0.020574986	K07133
K03540	1.6	1.08	0.000604131	0.020574986	RPR2, RPP21
K02015	3.9	2.64	0.000739776	0.024494794	ABC.FEV.P
K01155	0.6	0.12	0.001052513	0.033907979	E3.1.21.4
K00123	2.3	1.08	0.001093925	0.034314699	fdoG, fdhF, fdwA

K12994	0.5	0.08	0.001183422	0.035168987	wbdB, wbpY
K02018	1.9	0.96	0.001238594	0.035168987	modB
K07335	0.9	0.32	0.001199546	0.035168987	bmpA, bmpB, tmpC
K19294	0.9	0.28	0.001239176	0.035168987	algI
K00058	1.8	1.08	0.001524552	0.042262009	serA, PHGDH
K00125	2.6	1.72	0.001769156	0.047928042	fdhB

Table 6.22. KEGG Orthologs (KO) enriched in wombat and mahogany glider *Methanocorpusculum* genomes. KO annotation and statistical analysis were performed using 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). Only KOs with p values of <0.05 were retained and considered significant, as determined by Fisher's Exact Test. Six wombat and five mahogany glider genomes were included in the analysis.

KO	Wombat Gen.	Glider Gen.	P-value	Corrected P-value	Name	Description
K04719	0	5	0.002164502	0.047127115	bluB	5,6-dimethylbenzimidazole synthase [EC:1.13.11.79]
K11751	5	0	0.015151515	0.216644052	ushA	5'-nucleotidase / UDP-sugar diphosphatase [EC:3.1.3.5 3.6.1.45]
K06897	6	0	0.002164502	0.047127115	K06897	7,8-dihydropterin-6-yl-methyl-4-(beta-D-ribofuranosyl)aminobenzene 5'-phosphate synthase [EC:2.5.1.105]
K01652	0	4	0.015151515	0.216644052	ilvB, ilvG, ilvI	acetolactate synthase I/II/III large subunit [EC:2.2.1.6]
K01653	0	4	0.015151515	0.216644052	ilvH, ilvN	acetolactate synthase I/III small subunit [EC:2.2.1.6]
K03310	0	5	0.002164502	0.047127115	TC.AGCS	alanine or glycine:cation symporter, AGCS family
K19954	5	0	0.015151515	0.216644052	adh1	alcohol dehydrogenase [EC:1.1.1.-]
K12994	5	0	0.015151515	0.216644052	wbdB, wbpY	alpha-1,3-rhamnosyltransferase [EC:2.4.1.349 2.4.1.-]
K15770	0	5	0.002164502	0.047127115	cycB, ganO	arabinogalactan oligomer / maltooligosaccharide transport system substrate-binding protein
K06147	0	5	0.002164502	0.047127115	ABCB-BAC	ATP-binding cassette, subfamily B, bacterial
K03655	0	5	0.002164502	0.047127115	recG	ATP-dependent DNA helicase RecG [EC:3.6.4.12]
K01709	0	5	0.002164502	0.047127115	rfbG	CDP-glucose 4,6-dehydratase [EC:4.2.1.45]
K12710	0	5	0.002164502	0.047127115	novU	C-methyltransferase [EC:2.1.1.-]
K03183	6	0	0.002164502	0.047127115	ubiE	demethylmenaquinone methyltransferase [EC:2.1.1.163 2.1.1.201]
K00382	6	1	0.015151515	0.216644052	DLD, lpd, pdhD	dihydrolipoamide dehydrogenase [EC:1.8.1.4]
K01687	0	5	0.002164502	0.047127115	ilvD	dihydroxy-acid dehydratase [EC:4.2.1.9]
K07343	0	5	0.002164502	0.047127115	tfoX	DNA transformation protein and related proteins
K01790	0	5	0.002164502	0.047127115	rfbC, rmlC	dTDP-4-dehydrorhamnose 3,5-epimerase [EC:5.1.3.13]
K19180	0	5	0.002164502	0.047127115	tll	dTDP-6-deoxy-L-talose 4-dehydrogenase (NAD+) [EC:1.1.1.339]
K00299	0	4	0.015151515	0.216644052	ssuE	FMN reductase [EC:1.5.1.38]

K10563	6	1	0.015151515	0.216644052	mutM, fpg	formamidopyrimidine-DNA glycosylase [EC:3.2.2.23 4.2.99.18]
K00203	6	1	0.015151515	0.216644052	fwdD, fmdD	formylmethanofuran dehydrogenase subunit D [EC:1.2.7.12]
K01678	6	1	0.015151515	0.216644052	fumB	fumarate hydratase subunit beta [EC:4.2.1.2]
K00978	0	5	0.002164502	0.047127115	rfbF	glucose-1-phosphate cytidylyltransferase [EC:2.7.7.33]
K11130	1	5	0.015151515	0.216644052	NOP10, NOLA3	H/ACA ribonucleoprotein complex subunit 3
K01546	6	0	0.002164502	0.047127115	kdpA	K+-transporting ATPase ATPase A chain [EC:3.6.3.12]
K01547	6	0	0.002164502	0.047127115	kdpB	K+-transporting ATPase ATPase B chain [EC:3.6.3.12]
K18009	6	0	0.002164502	0.047127115	budC	meso-butanediol dehydrogenase [EC:1.1.1.- 1.1.1.76 1.1.1.304]
K10977	0	5	0.002164502	0.047127115	aksA	methanogen homocitrate synthase [EC:2.3.3.14 2.3.3.-]
K03315	0	5	0.002164502	0.047127115	nhaC	Na+:H+ antiporter, NhaC family
K02589	0	5	0.002164502	0.047127115	nifHD1, nifI1	nitrogen regulatory protein PII 1
K02590	0	5	0.002164502	0.047127115	nifHD2, nifI2	nitrogen regulatory protein PII 2
K02587	0	5	0.002164502	0.047127115	nifE	nitrogenase molybdenum-cofactor synthesis protein NifE
K02586	0	5	0.002164502	0.047127115	nifD	nitrogenase molybdenum-iron protein alpha chain [EC:1.18.6.1]
K02591	0	5	0.002164502	0.047127115	nifK	nitrogenase molybdenum-iron protein beta chain [EC:1.18.6.1]
K02592	0	5	0.002164502	0.047127115	nifN	nitrogenase molybdenum-iron protein NifN
K15583	6	0	0.002164502	0.047127115	oppD	oligopeptide transport system ATP-binding protein
K10823	6	0	0.002164502	0.047127115	oppF	oligopeptide transport system ATP-binding protein
K02031	0	5	0.002164502	0.047127115	ABC.PE.A	peptide/nickel transport system ATP-binding protein
K02032	0	5	0.002164502	0.047127115	ABC.PE.A1	peptide/nickel transport system ATP-binding protein
K13010	0	5	0.002164502	0.047127115	per, rfbE	perosamine synthetase [EC:2.6.1.102]
K02040	6	0	0.002164502	0.047127115	pstS	phosphate transport system substrate-binding protein
K01841	6	1	0.015151515	0.216644052	pepM	phosphoenolpyruvate phosphomutase [EC:5.4.2.9]
K09459	6	1	0.015151515	0.216644052	E4.1.1.82	phosphonopyruvate decarboxylase [EC:4.1.1.82]
K01496	0	5	0.002164502	0.047127115	hisI	phosphoribosyl-AMP cyclohydrolase [EC:3.5.4.19]

K11626	6	0	0.002164502	0.047127115	glnT	putative sodium/glutamine symporter
K07497	5	0	0.015151515	0.216644052	K07497	putative transposase
K13280	0	5	0.002164502	0.047127115	SEC11, sipW	signal peptidase I [EC:3.4.21.89]
K16868	5	0	0.015151515	0.216644052	tehB	tellurite methyltransferase [EC:2.1.1.265]
K18218	6	0	0.002164502	0.047127115	tet35	tetracycline resistance efflux pump
K03672	6	0	0.002164502	0.047127115	trxC	thioredoxin 2 [EC:1.8.1.8]
K01754	0	4	0.015151515	0.216644052	ilvA, tdcB	threonine dehydratase [EC:4.3.1.19]
K15449	6	1	0.015151515	0.216644052	TYW1	tRNA wybutosine-synthesizing protein 1 [EC:4.1.3.44]
K06857	6	1	0.015151515	0.216644052	tupC, vupC	tungstate transport system ATP-binding protein [EC:3.6.3.55]
K05773	6	1	0.015151515	0.216644052	tupB, vupB	tungstate transport system permease protein
K05772	6	1	0.015151515	0.216644052	tupA, vupA	tungstate transport system substrate-binding protein
K07646	6	0	0.002164502	0.047127115	kdpD	two-component system, OmpR family, sensor histidine kinase KdpD [EC:2.7.13.3]
K01155	6	0	0.002164502	0.047127115	E3.1.21.4	type II restriction enzyme [EC:3.1.21.4]
K08678	0	4	0.015151515	0.216644052	UXS1, uxs	UDP-glucuronate decarboxylase [EC:4.1.1.35]
K06888	0	4	0.015151515	0.216644052	K06888	uncharacterized protein
K06921	0	5	0.002164502	0.047127115	K06921	uncharacterized protein
K06923	0	4	0.015151515	0.216644052	K06923	uncharacterized protein
K07124	6	0	0.002164502	0.047127115	K07124	uncharacterized protein
K09815	0	5	0.002164502	0.047127115	znuA	zinc transport system substrate-binding protein
K02074	6	0	0.002164502	0.047127115	ABC.ZM.A	zinc/manganese transport system ATP-binding protein
K02075	6	0	0.002164502	0.047127115	ABC.ZM.P	zinc/manganese transport system permease protein
K02077	6	0	0.002164502	0.047127115	ABC.ZM.S	zinc/manganese transport system substrate-binding protein

Table 6.23. KEGG Orthologs (KO) enriched in wombat and mahogany glider *Methanocorpusculum* genomes. KO annotation and statistical analysis were performed using 'annotate' and 'enrichment' functions of EnrichM (v0.4.9). Only KOs with p values of <0.05 were retained and considered significant, as determined by Mann-Whitney U Test. Six wombat and five mahogany glider genomes were included in the analysis.

KO	Wombat Gene Count	Glider Gene Count	P Value	Corrected P Value	Gene
K00200	3.833333333	2.2	0.002910433	1	fwdA, fmdA
K00202	3.833333333	2.2	0.002910433	1	fwdC, fmdC
K00201	3.833333333	1.8	0.003455206	1	fwdB, fmdB
K00123	2.833333333	1.4	0.012168349	1	fdoG, fdhF, fdwA
K07088	0.833333333	1.6	0.020984445	1	K07088
K01673	1.5	0.4	0.026538443	1	cynT, can
K01952	2	1.2	0.027586466	1	purL, PFAS
K01711	0.833333333	0.2	0.028889786	1	gmd, GMDS
K02377	0.833333333	0.2	0.028889786	1	TSTA3, fcl
K09714	1.833333333	1.2	0.028889786	1	K09714
K03750	1.833333333	1.2	0.028889786	1	moeA
K02435	0.166666667	0.8	0.028889786	1	gatC, GATC
K08169	3.166666667	3.8	0.028889786	1	yebQ

Chapter 7: References

Abu-Qarn, M., & Eichler, J. (2007). An analysis of amino acid sequences surrounding archaeal glycoprotein sequons. *Archaea*, *2*(2), 73-81. https://doi.org/10.1155/2006/510578

Advani, V. M., & Dinman, J. D. (2016). Reprogramming the genetic code: The emerging role of ribosomal frameshifting in regulating cellular gene expression. *Bioessays*, *38*(1), 21-26. https://doi.org/10.1002/bies.201500131

Akileswaran, L., Brock, B. J., Cereghino, J. L., & Gold, M. H. (1999). 1,4-benzoquinone reductase from Phanerochaete chrysosporium: cDNA cloning and regulation of expression. *Appl Environ Microbiol*, *65*(2), 415-421. https://doi.org/10.1128/aem.65.2.415-421.1999

Al-Masaudi, S., El Kaoutari, A., Drula, E., Al-Mehdar, H., Redwan, E. M., Lombard, V., & Henrissat, B. (2017). A Metagenomics Investigation of Carbohydrate-Active Enzymes along the Gastrointestinal Tract of Saudi Sheep. *Front Microbiol*, *8*, 666. https://doi.org/10.3389/fmicb.2017.00666

Almeida, A., Mitchell, A. L., Boland, M., Forster, S. C., Gloor, G. B., Tarkowska, A., Lawley, T. D., & Finn, R. D. (2019). A new genomic blueprint of the human gut microbiota. *Nature*, *568*(7753), 499-504. https://doi.org/10.1038/s41586-019-0965-1

Alvarez-Martinez, C. E., & Christie, P. J. (2009). Biological Diversity of Prokaryotic Type IV Secretion Systems. *Microbiology and Molecular Biology Reviews*, *73*(4), 775-808. https://doi.org/10.1128/mmbr.00023-09

Anderson, I., Ulrich, L. E., Lupa, B., Susanti, D., Porat, I., Hooper, S. D., Lykidis, A., Sieprawska-Lupa, M., Dharmarajan, L., Goltsman, E., Lapidus, A., Saunders, E., Han, C., Land, M., Lucas, S., Mukhopadhyay, B., Whitman, W. B., Woese, C., Bristow, J., & Kyrpides, N. (2009). Genomic characterization of methanomicrobiales reveals three classes of methanogens. *Plos One*, *4*(6), e5797. https://doi.org/10.1371/journal.pone.0005797

Ang, A., Roesma, D. I., Nijman, V., Meier, R., Srivathsan, A., & Rizaldi. (2020). Faecal DNA to the rescue: Shotgun sequencing of non-invasive samples reveals two subspecies of Southeast Asian primates to be Critically Endangered species. *Sci Rep*, *10*(1), 9396. https://doi.org/10.1038/s41598-020-66007-8

Antonov, I., Coakley, A., Atkins, J. F., Baranov, P. V., & Borodovsky, M. (2013). Identification of the nature of reading frame transitions observed in prokaryotic genomes. *Nucleic Acids Res*, *41*(13), 6514-6530. https://doi.org/10.1093/nar/gkt274

Araos, R., Battaglia, T., Ugalde, J. A., Rojas-Herrera, M., Blaser, M. J., & D'Agata, E. M. C. (2019). Fecal Microbiome Characteristics and the Resistome Associated With Acquisition

of Multidrug-Resistant Organisms Among Elderly Subjects. *Front Microbiol*, *10*, 2260. https://doi.org/10.3389/fmicb.2019.02260

Armougom, F., Henry, M., Vialettes, B., Raccah, D., & Raoult, D. (2009). Monitoring bacterial community of human gut microbiota reveals an increase in Lactobacillus in obese patients and Methanogens in anorexic patients. *Plos One*, *4*(9), e7125. https://doi.org/10.1371/journal.pone.0007125

Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E., & Baranov, P. V. (2016). Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res*, *44*(15), 7007-7078. https://doi.org/10.1093/nar/gkw530

Attaluri, A., Jackson, M., Valestin, J., & Rao, S. S. (2010). Methanogenic flora is associated with altered colonic transit but not stool characteristics in constipation without IBS. *Am J Gastroenterol*, *105*(6), 1407-1411. https://doi.org/10.1038/ajg.2009.655

Auffret, M. D., Stewart, R., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., Wallace, R. J., Freeman, T. C., Snelling, T. J., Watson, M., & Roehe, R. (2017). Identification, Comparison, and Validation of Robust Rumen Microbial Biomarkers for Methane Emissions Using Diverse Bos Taurus Breeds and Basal Diets. *Front Microbiol*, *8*, 2642. https://doi.org/10.3389/fmicb.2017.02642

Ayesh, R., Mitchell, S. C., Zhang, A., & Smith, R. L. (1993). The fish odour syndrome: biochemical, familial, and clinical aspects. *BMJ : British Medical Journal*, *307*(6905), 655-657. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1678986/

Balch, W. E., Fox, G. E., Magrum, L. J., Woese, C. R., & Wolfe, R. S. (1979). Methanogens: reevaluation of a unique biological group. *Microbiol Rev*, *43*(2), 260-296. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC281474/pdf/microrev00006-0148.pdf

Bang, C., Schilhabel, A., Weidenbach, K., Kopp, A., Goldmann, T., Gutsmann, T., & Schmitz, R. A. (2012). Effects of antimicrobial peptides on methanogenic archaea. *Antimicrob Agents Chemother*, *56*(8), 4123-4130. https://doi.org/10.1128/aac.00661-12

Bang, C., & Schmitz, R. A. (2015). Archaea associated with human surfaces: not to be underestimated. *FEMS Microbiol Rev*, *39*(5), 631-648. https://doi.org/10.1093/femsre/fuv010

Bang, C., Vierbuchen, T., Gutsmann, T., Heine, H., & Schmitz, R. A. (2017). Immunogenic properties of the human gut-associated archaeon Methanomassiliicoccus luminyensis and its susceptibility to antimicrobial peptides. *Plos One*, *12*(10), e0185919-e0185919. https://doi.org/10.1371/journal.pone.0185919

Bang, C., Weidenbach, K., Gutsmann, T., Heine, H., & Schmitz, R. A. (2014). The intestinal archaea Methanosphaera stadtmanae and Methanobrevibacter smithii activate human dendritic cells. *Plos One*, *9*(6), e99411. https://doi.org/10.1371/journal.pone.0099411

Bapteste, E., Brochier, C., & Boucher, Y. (2005). Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea*, *1*(5), 353-363. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2685549/pdf/Archaea-01-353.pdf

Barco, R. A., Garrity, G. M., Scott, J. J., Amend, J. P., Nealson, K. H., & Emerson, D. (2020). A Genus Definition for Bacteria and Archaea Based on a Standard Genome Relatedness Index. *mBio*, *11*(1), e02475-02419. https://doi.org/10.1128/mBio.02475-19

Barker, C. J., Gillett, A., Polkinghorne, A., & Timms, P. (2013). Investigation of the koala (Phascolarctos cinereus) hindgut microbiome via 16S pyrosequencing. *Veterinary Microbiology*, *167*(3), 554-564. https://doi.org/https://doi.org/10.1016/j.vetmic.2013.08.025

Barrasa, J. I., Olmo, N., Lizarbe, M. A., & Turnay, J. (2013). Bile acids in the colon, from healthy to cytotoxic molecules. *Toxicol In Vitro*, *27*(2), 964-977. https://doi.org/10.1016/j.tiv.2012.12.020

Barret, M., Gagnon, N., Kalmokoff, M. L., Topp, E., Verastegui, Y., Brooks, S. P., Matias, F., Neufeld, J. D., & Talbot, G. (2013). Identification of Methanoculleus spp. as active methanogens during anoxic incubations of swine manure storage tank samples. *Appl Environ Microbiol*, *79*(2), 424-433. https://doi.org/10.1128/aem.02268-12

Beck, R. M. D., Louys, J., Brewer, P., Archer, M., Black, K. H., & Tedford, R. H. (2020). A new family of diprotodontian marsupials from the latest Oligocene of Australia and the evolution of wombats, koalas, and their relatives (Vombatiformes). *Sci Rep*, *10*(1), 9741. https://doi.org/10.1038/s41598-020-66425-8

Bedarf, J. R., Hildebrand, F., Coelho, L. P., Sunagawa, S., Bahram, M., Goeser, F., Bork, P., & Wüllner, U. (2017). Functional implications of microbial and viral gut metagenome changes in early stage L-DOPA-naïve Parkinson's disease patients. *Genome Med*, *9*(1), 39. https://doi.org/10.1186/s13073-017-0428-y

Belay, N., Johnson, R., Rajagopal, B. S., Conway de Macario, E., & Daniels, L. (1988). Methanogenic bacteria from human dental plaque. *Appl Environ Microbiol*, *54*(2), 600-603. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC202503/pdf/aem00107-0328.pdf

Belay, N., Mukhopadhyay, B., Conway de Macario, E., Galask, R., & Daniels, L. (1990). Methanogenic bacteria in human vaginal samples. *Journal of Clinical Microbiology*, *28*(7), 1666. http://jcm.asm.org/content/28/7/1666.abstract

Belkacemi, S., Mazel, A., Tardivo, D., Tavitian, P., Stephan, G., Bianca, G., Terrer, E., Drancourt, M., & Aboudharam, G. (2018). Peri-implantitis-associated methanogens: a preliminary report. *Sci Rep*, *8*(1), 9447. https://doi.org/10.1038/s41598-018-27862-8

Bernatchez, E., Gold, M. J., Langlois, A., Blais-Lecours, P., Boucher, M., Duchaine, C., Marsolais, D., McNagny, K. M., & Blanchet, M. R. (2017). Methanosphaera stadtmanae induces a type IV hypersensitivity response in a mouse model of airway inflammation. *Physiol Rep*, *5*(7). https://doi.org/10.14814/phy2.13163

Berndt, C., Lillig, C. H., Wollenberg, M., Bill, E., Mansilla, M. C., de Mendoza, D., Seidler, A., & Schwenn, J. D. (2004). Characterization and reconstitution of a 4Fe-4S adenylyl sulfate/phosphoadenylyl sulfate reductase from Bacillus subtilis. *J Biol Chem*, *279*(9), 7850-7855. https://doi.org/10.1074/jbc.M309332200

Biavati, B., Vasta, M., & Ferry, J. G. (1988). Isolation and characterization of "Methanosphaera cuniculi" sp. nov. *Appl Environ Microbiol*, *54*(3), 768-771. https://doi.org/10.1128/aem.54.3.768-771.1988

Black, K. H., Archer, M., Hand, S. J., & Godthelp, H. (2012). The Rise of Australian Marsupials: A Synopsis of Biostratigraphic, Phylogenetic, Palaeoecologic and Palaeobiogeographic Understanding. In J. A. Talent (Ed.), *Earth and Life: Global Biodiversity, Extinction Intervals and Biogeographic Perturbations Through Time* (pp. 983-1078). Springer Netherlands. https://doi.org/10.1007/978-90-481-3428-1_35

Blais Lecours, P., Duchaine, C., Taillefer, M., Tremblay, C., Veillette, M., Cormier, Y., & Marsolais, D. (2011). Immunogenic Properties of Archaeal Species Found in Bioaerosols. *Plos One*, *6*(8), Article e23326. https://doi.org/10.1371/journal.pone.0023326

Blais Lecours, P., Marsolais, D., Cormier, Y., Berberi, M., Hache, C., Bourdages, R., & Duchaine, C. (2014). Increased Prevalence of Methanosphaera stadtmanae in Inflammatory Bowel Diseases: e87734. *Plos One*, *9*(2). https://doi.org/10.1371/journal.pone.0087734

Bleicher, K., & Winter, J. (1991). Purification and properties of F420- and NADP(+)-dependent alcohol dehydrogenases of Methanogenium liminatans and Methanobacterium palustre, specific for secondary alcohols. *Eur J Biochem*, *200*(1), 43-51. https://doi.org/10.1111/j.1432-1033.1991.tb21046.x

Bleicher, K., Zellner, G., & Winter, J. (1989). Growth of methanogens on cyclopentanol/CO2 and specificity of alcohol dehydrogenase. *Fems Microbiology Letters*, *59*(3), 307-312. https://doi.org/10.1111/j.1574-6968.1989.tb03130.x

Boccia, S., Torre, I., Santarpia, L., Iervolino, C., Del Piano, C., Puggina, A., Pastorino, R., Dragic, M., Amore, R., Borriello, T., Palladino, R., Pennino, F., Contaldo, F., & Pasanisi, F. (2017).

Intestinal microbiota in adult patients with Short Bowel Syndrome: Preliminary results from a pilot study. *Clin Nutr*, *36*(6), 1707-1709. https://doi.org/10.1016/j.clnu.2016.09.028

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Bordoni, L., Samulak, J. J., Sawicka, A. K., Pelikant-Malecka, I., Radulska, A., Lewicki, L., Kalinowski, L., Gabbianelli, R., & Olek, R. A. (2020). Trimethylamine N-oxide and the reverse cholesterol transport in cardiovascular disease: a cross-sectional study. *Sci Rep*, *10*(1), 18675. https://doi.org/10.1038/s41598-020-75633-1

Borgo, F., Riva, A., Benetti, A., Casiraghi, M. C., Bertelli, S., Garbossa, S., Anselmetti, S., Scarone, S., Pontiroli, A. E., Morace, G., & Borghi, E. (2017). Microbiota in anorexia nervosa: The triangle between bacterial species, metabolites and psychological tests. *Plos One*, *12*(6), e0179739. https://doi.org/10.1371/journal.pone.0179739

Borrel, G., Harris, H. M., Parisot, N., Gaci, N., Tottey, W., Mihajlovski, A., Deane, J., Gribaldo, S., Bardot, O., Peyretaillade, E., Peyret, P., O'Toole, P. W., & Brugere, J. F. (2013). Genome Sequence of "Candidatus Methanomassiliicoccus intestinalis" Issoire-Mx1, a Third Thermoplasmatales-Related Methanogenic Archaeon from Human Feces. *Genome Announc*, *1*(4). https://doi.org/10.1128/genomeA.00453-13

Borrel, G., Harris, H. M., Tottey, W., Mihajlovski, A., Parisot, N., Peyretaillade, E., Peyret, P., Gribaldo, S., O'Toole, P. W., & Brugere, J. F. (2012). Genome sequence of "Candidatus Methanomethylophilus alvus" Mx1201, a methanogenic archaeon from the human gut belonging to a seventh order of methanogens. *J Bacteriol*, *194*(24), 6944-6945. https://doi.org/10.1128/jb.01867-12

Borrel, G., O'Toole, P. W., Harris, H. M., Peyret, P., Brugere, J. F., & Gribaldo, S. (2013). Phylogenomic data support a seventh order of Methylotrophic methanogens and provide insights into the evolution of Methanogenesis. *Genome Biol Evol*, *5*(10), 1769-1780. https://doi.org/10.1093/gbe/evt128

Borrel, G., Parisot, N., Harris, H. M., Peyretaillade, E., Gaci, N., Tottey, W., Bardot, O., Raymann, K., Gribaldo, S., Peyret, P., O'Toole, P. W., & Brugere, J. F. (2014). Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. *BMC Genomics*, *15*, 679. https://doi.org/10.1186/1471-2164-15-679

Boyd, E. S., Anbar, A. D., Miller, S., Hamilton, T. L., Lavin, M., & Peters, J. W. (2011). A late methanogen origin for molybdenum-dependent nitrogenase. *Geobiology*, *9*(3), 221-232. https://doi.org/10.1111/j.1472-4669.2011.00278.x

Boyd, J. A., Woodcroft, B. J., & Tyson, G. W. (2018). GraftM: a tool for scalable, phylogenetically informed classification of genes within metagenomes. *Nucleic Acids Res*, *46*(10), e59-e59. https://doi.org/10.1093/nar/gky174

Boyd, J. A., Woodcroft, B. J., & Tyson, G. W. (2019). Comparative genomics using EnrichM. . (In preparation).

Boyer, J. L. (2013). Bile formation and secretion. *Comprehensive Physiology*, *3*(3), 1035-1078. https://doi.org/10.1002/cphy.c120027

Bratten, J. R., Spanier, J., & Jones, M. P. (2008). Lactulose breath testing does not discriminate patients with irritable bowel syndrome from healthy controls. *Am J Gastroenterol*, *103*(4), 958-963. https://doi.org/10.1111/j.1572-0241.2008.01785.x

Brice, K. L., Trivedi, P., Jeffries, T. C., Blyton, M. D. J., Mitchell, C., Singh, B. K., & Moore, B. D. (2019). The Koala (Phascolarctos cinereus) faecal microbiome differs with diet in a wild population. *PeerJ*, *7*, e6534. https://doi.org/10.7717/peerj.6534

Brito, I. L., Gurry, T., Zhao, S., Huang, K., Young, S. K., Shea, T. P., Naisilisili, W., Jenkins, A. P., Jupiter, S. D., Gevers, D., & Alm, E. J. (2019). Transmission of human-associated microbiota along family and social networks. *Nat Microbiol*, *4*(6), 964-971. https://doi.org/10.1038/s41564-019-0409-6

Brugere, J. F., Borrel, G., Gaci, N., Tottey, W., O'Toole, P. W., & Malpuech-Brugere, C. (2014). Archaebiotics: proposed therapeutic use of archaea to prevent trimethylaminuria and cardiovascular disease. *Gut Microbes*, *5*(1), 5-10. https://doi.org/10.4161/gmic.26749

Burén, S., Pratt, K., Jiang, X., Guo, Y., Jimenez-Vicente, E., Echavarri-Erasun, C., Dean, D. R., Saaem, I., Gordon, D. B., Voigt, C. A., & Rubio, L. M. (2019). Biosynthesis of the nitrogenase active-site cofactor precursor NifB-co in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, *116*(50), 25078-25086. https://doi.org/10.1073/pnas.1904903116

Burke, C., Burnard, D., Polkinghorne, A., Webb, J., & Huston, W. M. (2018). Cloacal and Ocular Microbiota of the Endangered Australian Northern Quoll. *Microorganisms*, *6*(3). https://doi.org/10.3390/microorganisms6030068

Camara, A., Konate, S., Tidjani Alou, M., Kodio, A., Togo, A. H., Cortaredona, S., Henrissat, B., Thera, M. A., Doumbo, O. K., Raoult, D., & Million, M. (2021). Clinical evidence of the role of Methanobrevibacter smithii in severe acute malnutrition. *Sci Rep*, *11*(1), 5426. https://doi.org/10.1038/s41598-021-84641-8

Campbell, T. P., Sun, X., Patel, V. H., Sanz, C., Morgan, D., & Dantas, G. (2020). The microbiome and resistome of chimpanzees, gorillas, and humans across host lifestyle and geography. *Isme j*, *14*(6), 1584-1599. https://doi.org/10.1038/s41396-020-0634-2

Cao, J., Hu, Y., Liu, F., Wang, Y., Bi, Y., Lv, N., Li, J., Zhu, B., & Gao, G. F. (2020). Metagenomic analysis reveals the microbiome and resistome in migratory birds. *Microbiome*, *8*(1), 26. https://doi.org/10.1186/s40168-019-0781-8

Cao, Y., Xu, H., Li, R., Gao, S., Chen, N., Luo, J., & Jiang, Y. (2019). Genetic Basis of Phenotypic Differences Between Chinese Yunling Black Goats and Nubian Goats Revealed by Allele-Specific Expression in Their F1 Hybrids. *Front Genet*, *10*, 145. https://doi.org/10.3389/fgene.2019.00145

Carey, H. V., Walters, W. A., & Knight, R. (2013). Seasonal restructuring of the ground squirrel gut microbiota over the annual hibernation cycle. *American journal of physiology. Regulatory, integrative and comparative physiology*, *304*(1), R33-R42. https://doi.org/10.1152/ajpregu.00387.2012

Casey, F. F., Stannard, H. J., & Old, J. M. (2021). A review of wombat diet and nutrition. *Australian Mammalogy*, *43*(1), 1-9. https://doi.org/https://doi.org/10.1071/AM20009

Castillo-Alvarez, F., Perez-Matute, P., Oteo, J. A., & Marzo-Sola, M. E. (2018). The influence of interferon beta-1b on gut microbiota composition in patients with multiple sclerosis. *Neurologia*. https://doi.org/10.1016/j.nrl.2018.04.006 (Composicion de la microbiota intestinal en pacientes con esclerosis multiple. Influencia del tratamiento con interferonbeta-1b.)

Chaokaur, A., Nishida, T., Phaowphaisal, I., & Sommart, K. (2015). Effects of feeding level on methane emissions and energy utilization of Brahman cattle in the tropics. *Agriculture, Ecosystems & Environment*, *199*, 225-230. https://doi.org/https://doi.org/10.1016/j.agee.2014.09.014

Chatterjee, S., Park, S., Low, K., Kong, Y., & Pimentel, M. (2007). The degree of breath methane production in IBS correlates with the severity of constipation. *Am J Gastroenterol*, *102*(4), 837-841. https://doi.org/10.1111/j.1572-0241.2007.01072.x

Chaudhari, N. M., Gupta, V. K., & Dutta, C. (2016). BPGA- an ultra-fast pan-genome analysis pipeline. *Sci Rep*, *6*(1), 24373. https://doi.org/10.1038/srep24373

Chaudhary, P. P., Gaci, N., Borrel, G., O'Toole, P. W., & Brugere, J. F. (2015). Molecular methods for studying methanogens of the human gastrointestinal tract: current status and future directions. *Appl Microbiol Biotechnol*, *99*(14), 5801-5815. https://doi.org/10.1007/s00253-015-6739-2

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, *36*(6), 1925-1927. https://doi.org/10.1093/bioinformatics/btz848

Cheeseman, P., Toms-Wood, A., & Wolfe, R. S. (1972). Isolation and properties of a fluorescent compound, factor 420 , from Methanobacterium strain M.o.H. *J Bacteriol*, *112*(1), 527-531. http://jb.asm.org/content/112/1/527.full.pdf

Chekabab, S. M., Lawrence, J. R., Alvarado, A., Predicala, B., & Korber, D. R. (2020). A health metadata-based management approach for comparative analysis of high-throughput genetic sequences for quantifying antimicrobial resistance reduction in Canadian hog barns. *Comput Struct Biotechnol J*, *18*, 2629-2638. https://doi.org/10.1016/j.csbj.2020.09.012

Chen, H., Mozzicafreddo, M., Pierella, E., Carletti, V., Piersanti, A., Ali, S. M., Ame, S. M., Wang, C., & Miceli, C. (2021). Dissection of the gut microbiota in mothers and children with chronic Trichuris trichiura infection in Pemba Island, Tanzania. *Parasit Vectors*, *14*(1), 62. https://doi.org/10.1186/s13071-021-04580-1

Chen, I. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J. R., Seshadri, R., Smirnova, T., Kirton, E., Jungbluth, S. P., Woyke, T., Eloe-Fadrosh, E. A., Ivanova, N. N., & Kyrpides, N. C. (2019). IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res*, *47*(D1), D666-d677. https://doi.org/10.1093/nar/gky901

Chen, Q., Wang, J., Zhang, H., Shi, H., Liu, G., Che, J., & Liu, B. (2021). Microbial community and function in nitrogen transformation of ectopic fermentation bed system for pig manure composting. *Bioresource Technology*, *319*, 124155. https://doi.org/10.1016/j.biortech.2020.124155

Chen, Y.-M., Holmes, E. C., Chen, X., Tian, J.-H., Lin, X.-D., Qin, X.-C., Gao, W.-H., Liu, J., Wu, Z.-D., & Zhang, Y.-Z. (2020). Diverse and abundant resistome in terrestrial and aquatic vertebrates revealed by transcriptional analysis. *Sci Rep*, *10*(1), 18870. https://doi.org/10.1038/s41598-020-75904-x

Cheng, Y., Fox, S., Pemberton, D., Hogg, C., Papenfuss, A. T., & Belov, K. (2015). The Tasmanian devil microbiome-implications for conservation and management. *Microbiome*, *3*, 76. https://doi.org/10.1186/s40168-015-0143-0

Chiang, J. Y. (2017). Recent advances in understanding bile acid homeostasis. *F1000Res*, *6*, 2029. https://doi.org/10.12688/f1000research.12449.1

Chibani, C., Mahnert, A., Borrel, G., Almeida, A. W., A; , Brugère, J. G., S;, Finn, R., Schmitz, R., & Moissl-Eichinger, C. (2020). A Comprehensive Analysis of the Global Human Gut Archaeome from a Thousand Genome Catalogue. *Cell*.

Chojnacka, A., Szczęsny, P., Błaszczyk, M. K., Zielenkiewicz, U., Detman, A., Salamon, A., & Sikora, A. (2015). Noteworthy Facts about a Methane-Producing Microbial Community

Processing Acidic Effluent from Sugar Beet Molasses Fermentation. *Plos One*, *10*(5), e0128008. https://doi.org/10.1371/journal.pone.0128008

Christendat, D., Saridakis, V., Dharamsi, A., Bochkarev, A., Pai, E. F., Arrowsmith, C. H., & Edwards, A. M. (2000). Crystal structure of dTDP-4-keto-6-deoxy-D-hexulose 3,5-epimerase from Methanobacterium thermoautotrophicum complexed with dTDP. *J Biol Chem*, *275*(32), 24608-24612. https://doi.org/10.1074/jbc.C000238200

Chung, Y. W., Gwak, H. J., Moon, S., Rho, M., & Ryu, J. H. (2020). Functional dynamics of bacterial species in the mouse gut microbiome revealed by metagenomic and metatranscriptomic analyses. *Plos One*, *15*(1), e0227886. https://doi.org/10.1371/journal.pone.0227886

Cieslak, A., Szumacher-Strabel, M., Stochmal, A., & Oleszek, W. (2013). Plant components with specific activities against rumen methanogens. *Animal*, *7 Suppl 2*, 253-265. https://doi.org/10.1017/s1751731113000852

Clayton, J. B., Vangay, P., Huang, H., Ward, T., Hillmann, B. M., Al-Ghalith, G. A., Travis, D. A., Long, H. T., Tuan, B. V., Minh, V. V., Cabana, F., Nadler, T., Toddes, B., Murphy, T., Glander, K. E., Johnson, T. J., & Knights, D. (2016). Captivity humanizes the primate microbiome. *Proceedings of the National Academy of Sciences*, 201521835. https://doi.org/10.1073/pnas.1521835113

Coker, O. O., Wu, W. K. K., Wong, S. H., Sung, J. J. Y., & Yu, J. (2020). Altered Gut Archaea Composition and Interaction With Bacteria Are Associated With Colorectal Cancer. *Gastroenterology*, *159*(4), 1459-1470.e1455. https://doi.org/https://doi.org/10.1053/j.gastro.2020.06.042

Colombel, J.-F., Shin, A., & Gibson, P. R. (2018). Functional Gastrointestinal Symptoms in Patients With Inflammatory Bowel Disease. *Clinical Gastroenterology and Hepatology*. https://doi.org/https://doi.org/10.1016/j.cgh.2018.08.001

Cork, S. J., Hume, I. D., & Dawson, T. J. (1983). Digestion and metabolism of a natural foliar diet (Eucalyptus punctata) by an arboreal marsupial, the koala (Phascolarctos cinereus). *Journal of comparative physiology*, *153*(2), 181-190. https://doi.org/10.1007/BF00689622

Craciun, S., & Balskus, E. P. (2012). Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proceedings of the National Academy of Sciences*, *109*(52), 21307-21312. https://doi.org/10.1073/pnas.1215689109

D'Angelo, R., Esposito, T., Calabro, M., Rinaldi, C., Robledo, R., Varriale, B., & Sidoti, A. (2013). FMO3 allelic variants in Sicilian and Sardinian populations: trimethylaminuria and absence of fish-like body odor. *Gene*, *515*(2), 410-415. https://doi.org/10.1016/j.gene.2012.12.047

D'Arc, M., Furtado, C., Siqueira, J. D., Seuánez, H. N., Ayouba, A., Peeters, M., & Soares, M. A. (2018). Assessment of the gorilla gut virome in association with natural simian immunodeficiency virus infection. *Retrovirology*, *15*(1), 19. https://doi.org/10.1186/s12977-018-0402-9

Daddow, L. Y. M. (1983). A double lead stain method for enhancing contrast of ultrathin sections in electron microscopy: a modified multiple staining technique. *Journal of Microscopy*, *129*(2), 147-153. https://doi.org/https://doi.org/10.1111/j.1365-2818.1983.tb04169.x

Dahlhausen, K. E., Doroud, L., Firl, A. J., Polkinghorne, A., & Eisen, J. A. (2018). Characterization of shifts of koala (Phascolarctos cinereus) intestinal microbial communities associated with antibiotic treatment. *PeerJ*, *6*, e4452. https://doi.org/10.7717/peerj.4452

Dangal, S. R. S., Tian, H., Zhang, B., Pan, S., Lu, C., & Yang, J. (2017). Methane emission from global livestock sector during 1890–2014: Magnitude, trends and spatiotemporal patterns. *Global Change Biology*, *23*(10), 4147-4161. https://doi.org/10.1111/gcb.13709

Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *Plos One*, *5*(6), e11147. https://doi.org/10.1371/journal.pone.0011147

Das, R. C., & Heath, E. C. (1980). Dolichyldiphosphoryloligosaccharide--protein oligosaccharyltransferase; solubilization, purification, and properties. *Proc Natl Acad Sci U S A*, *77*(7), 3811-3815. https://doi.org/10.1073/pnas.77.7.3811

de Lacy Costello, B. P., Ledochowski, M., & Ratcliffe, N. M. (2013). The importance of methane breath testing: a review. *J Breath Res*, *7*(2), 024001. https://doi.org/10.1088/1752-7155/7/2/024001

DeLong, E. F. (1992). Archaea in coastal marine environments. *Proc Natl Acad Sci U S A*, *89*(12), 5685-5689. https://doi.org/10.1073/pnas.89.12.5685

Deppenmeier, U., & Muller, V. (2008). Life close to the thermodynamic limit: how methanogenic archaea conserve energy. *Results Probl Cell Differ*, *45*, 123-152. https://doi.org/10.1007/400_2006_026

Deusch, O., O'Flynn, C., Colyer, A., Swanson, K. S., Allaway, D., & Morris, P. (2015). A Longitudinal Study of the Feline Faecal Microbiome Identifies Changes into Early Adulthood Irrespective of Sexual Development. *Plos One*, *10*(12), e0144881. https://doi.org/10.1371/journal.pone.0144881

Dimidi, E., Christodoulides, S., Fragkos, K. C., Scott, S. M., & Whelan, K. (2014). The effect of probiotics on functional constipation in adults: a systematic review and meta-analysis of randomized controlled trials. *Am J Clin Nutr*, *100*(4), 1075-1084. https://doi.org/10.3945/ajcn.114.089151

Doden, H., Sallam, L. A., Devendran, S., Ly, L., Doden, G., Daniel, S. L., Alves, J. M. P., & Ridlon, J. M. (2018). Metabolism of Oxo-Bile Acids and Characterization of Recombinant 12alpha-Hydroxysteroid Dehydrogenases from Bile Acid 7alpha-Dehydroxylating Human Gut Bacteria. *Appl Environ Microbiol*, *84*(10). https://doi.org/10.1128/aem.00235-18

Dodt, W. G., Gallus, S., Phillips, M. J., & Nilsson, M. A. (2017). Resolving kangaroo phylogeny and overcoming retrotransposon ascertainment bias. *Sci Rep*, *7*(1), 16811. https://doi.org/10.1038/s41598-017-16148-0

Dongowski, G., Lorenz, A., & Anger, H. (2000). Degradation of pectins with different degrees of esterification by Bacteroides thetaiotaomicron isolated from human gut flora. *Appl Environ Microbiol*, *66*(4), 1321-1327. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC91987/pdf/am001321.pdf

Donovan, M., Lynch, M. D. J., Mackey, C. S., Platt, G. N., Washburn, B. K., Vera, D. L., Trickey, D. J., Charles, T. C., Wang, Z., & Jones, K. M. (2020). Metagenome-Assembled Genome Sequences of Five Strains from the Microtus ochrogaster (Prairie Vole) Fecal Microbiome. *Microbiol Resour Announc*, *9*(2). https://doi.org/10.1128/mra.01310-19

Doster, E., Rovira, P., Noyes, N. R., Burgess, B. A., Yang, X., Weinroth, M. D., Lakin, S. M., Dean, C. J., Linke, L., Magnuson, R., Jones, K. I., Boucher, C., Ruiz, J., Belk, K. E., & Morley, P. S. (2018). Investigating Effects of Tulathromycin Metaphylaxis on the Fecal Resistome and Microbiome of Commercial Feedlot Cattle Early in the Feeding Period. *Front Microbiol*, *9*, 1715. https://doi.org/10.3389/fmicb.2018.01715

Drancourt, M., Nkamga, V. D., Lakhe, N. A., Regis, J. M., Dufour, H., Fournier, P. E., Bechah, Y., Scheld, W. M., & Raoult, D. (2017). Evidence of Archaeal Methanogens in Brain Abscess. *Clin Infect Dis*, *65*(1), 1-5. https://doi.org/10.1093/cid/cix286

Dridi, B. (2012). Laboratory tools for detection of archaea in humans. *Clinical Microbiology and Infection*, *18*(9), 825-833. https://doi.org/10.1111/j.1469-0691.2012.03952.x

Dridi, B., Fardeau, M. L., Ollivier, B., Raoult, D., & Drancourt, M. (2011). The antimicrobial resistance pattern of cultured human methanogens reflects the unique phylogenetic position of archaea. *J Antimicrob Chemother*, *66*(9), 2038-2044. https://doi.org/10.1093/jac/dkr251

Dridi, B., Fardeau, M. L., Ollivier, B., Raoult, D., & Drancourt, M. (2012). Methanomassiliicoccus luminyensis gen. nov., sp nov., a methanogenic archaeon isolated from human faeces. *International Journal of Systematic and Evolutionary Microbiology*, *62*, 1902-1907. https://doi.org/10.1099/ijs.0.033712-0

Dridi, B., Henry, M., El Khechine, A., Raoult, D., & Drancourt, M. (2009). High Prevalence of Methanobrevibacter smithii and Methanosphaera stadtmanae Detected in the Human Gut

Using an Improved DNA Detection Protocol. *Plos One*, *4*(9), Article e7063. https://doi.org/10.1371/journal.pone.0007063

Dridi, B., Henry, M., Richet, H., Raoult, D., & Drancourt, M. (2012). Age-related prevalence of Methanomassiliicoccus luminyensis in the human gut microbiome. *Apmis*, *120*(10), 773-777. https://doi.org/10.1111/j.1600-0463.2012.02899.x

Dridi, B., Raoult, D., & Drancourt, M. (2011). Archaea as emerging organisms in complex human microbiomes. *Anaerobe*, *17*(2), 56-63. https://doi.org/10.1016/j.anaerobe.2011.03.001

Duboc, H., Rajca, S., Rainteau, D., Benarous, D., Maubert, M. A., Quervain, E., Thomas, G., Barbu, V., Humbert, L., Despras, G., Bridonneau, C., Dumetz, F., Grill, J. P., Masliah, J., Beaugerie, L., Cosnes, J., Chazouilleres, O., Poupon, R., Wolf, C., Mallet, J. M., Langella, P., Trugnan, G., Sokol, H., & Seksik, P. (2013). Connecting dysbiosis, bile-acid dysmetabolism and gut inflammation in inflammatory bowel diseases. *Gut*, *62*(4), 531-539. https://doi.org/10.1136/gutjnl-2012-302578

Eberhard, I., Mcnamara, J., Pearse, R., & Southwell, I. (1975). Ingestion and Excretion of Eucalyptus Punctata D.c. And Its Essential Oil by the Koala, Phascolarctos Cinereus (Goldfuss). *Australian Journal of Zoology*, *23*(2), 169-179. https://doi.org/https://doi.org/10.1071/ZO9750169

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., & Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science*, *308*(5728), 1635-1638. https://doi.org/10.1126/science.1110591

Eichler, J. (2003). Facing extremes: archaeal surface-layer (glyco)proteins. *Microbiology (Reading)*, *149*(Pt 12), 3347-3351. https://doi.org/10.1099/mic.0.26591-0

Einsle, O., Tezcan, F. A., Andrade, S. L. A., Schmid, B., Yoshida, M., Howard, J. B., & Rees, D. C. (2002). Nitrogenase MoFe-Protein at 1.16 Å Resolution: A Central Ligand in the FeMo-Cofactor. *Science*, *297*(5587), 1696-1700. https://doi.org/10.1126/science.1073877

Eisenhofer, R., Helgen, K. M., & Taggart, D. (2021). Signatures of landscape and captivity in the gut microbiota of Southern Hairy-nosed Wombats (Lasiorhinus latifrons). *Anim Microbiome*, *3*(1), 4. https://doi.org/10.1186/s42523-020-00068-y

Engelbrektson, A., Kunin, V., Wrighton, K. C., Zvenigorodsky, N., Chen, F., Ochman, H., & Hugenholtz, P. (2010). Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *The ISME Journal*, *4*(5), 642-647. https://doi.org/10.1038/ismej.2009.153

Enticknap, J. J., Kelly, M., Peraud, O., & Hill, R. T. (2006). Characterization of a Culturable Alphaproteobacterial Symbiont Common to Many Marine Sponges and Evidence for

Vertical Transmission via Sponge Larvae. *Applied and Environmental Microbiology*, *72*(5), 3724. https://doi.org/10.1128/AEM.72.5.3724-3732.2006

Eschler, B. M., Pass, D. M., Willis, R., & Foley, W. J. (2000). Distribution of foliar formylated phloroglucinol derivatives amongst Eucalyptus species. *Biochem Syst Ecol*, *28*(9), 813-824.

Evans, P. (2011). *Culture- and Molecular-based Studies of the Archaeal communities Present in the Forestomach Microbiomes of the Tammar Wallaby (Macropus eugenii) and Western Grey Kangaroo (Macropus fuliginosus)* The University of Queensland].

Evans, P. N., Hinds, L. A., Sly, L. I., McSweeney, C. S., Morrison, M., & Wright, A.-D. G. (2009). Community Composition and Density of Methanogens in the Foregut of the Tammar Wallaby (Macropus eugenii). *Applied and Environmental Microbiology*, *75*(8), 2598. https://doi.org/10.1128/AEM.02436-08

Eyssen, H. J., De Pauw, G., & Van Eldere, J. (1999). Formation of hyodeoxycholic acid from muricholic acid and hyocholic acid by an unidentified gram-positive rod termed HDCA-1 isolated from rat intestinal microflora. *Applied and Environmental Microbiology*, *65*(7), 3158-3163. https://doi.org/10.1128/AEM.65.7.3158-3163.1999

Fan, Y., Evans, C. R., Barber, K. W., Banerjee, K., Weiss, K. J., Margolin, W., Igoshin, O. A., Rinehart, J., & Ling, J. (2017). Heterogeneity of Stop Codon Readthrough in Single Bacterial Cells and Implications for Population Fitness. *Mol Cell*, *67*(5), 826-836.e825. https://doi.org/10.1016/j.molcel.2017.07.010

Fang, F., Li, Y., Bumann, M., Raftis, E. J., Casey, P. G., Cooney, J. C., Walsh, M. A., & O'Toole, P. W. (2009). Allelic variation of bile salt hydrolase genes in Lactobacillus salivarius does not determine bile resistance levels. *J Bacteriol*, *191*(18), 5743-5757. https://doi.org/10.1128/jb.00506-09

Fanzo, J., Covic, N., Dobermann, A., Henson, S., Herrero, M., Pingali, P., & Staal, S. (2020). A research vision for food systems in the 2020s: Defying the status quo. *Glob Food Sec*, *26*, 100397. https://doi.org/10.1016/j.gfs.2020.100397

Fernandes, J., Wang, A., Su, W., Rozenbloom, S. R., Taibi, A., Comelli, E. M., & Wolever, T. M. (2013). Age, dietary fiber, breath methane, and fecal short chain fatty acids are interrelated in Archaea-positive humans. *J Nutr*, *143*(8), 1269-1275. https://doi.org/10.3945/jn.112.170894

Fernandes, K. A., Kittelmann, S., Rogers, C. W., Gee, E. K., Bolwell, C. F., Bermingham, E. N., & Thomas, D. G. (2014). Faecal microbiota of forage-fed horses in New Zealand and the population dynamics of microbial communities following dietary change. *Plos One*, *9*(11), e112846. https://doi.org/10.1371/journal.pone.0112846

Ferrari, A., Brusa, T., Rutili, A., Canzi, E., & Biavati, B. (1994). Isolation and characterization ofMethanobrevibacter oralis sp. nov. *Curr Microbiol*, *29*(1), 7-12. https://doi.org/10.1007/BF01570184

Ferry, J. G. (1997). Enzymology of the fermentation of acetate to methane by Methanosarcina thermophila. *Biofactors*, *6*(1), 25-35. https://iubmb.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/biof.5520060104?download=true

Ferry, J. G. (1999). Enzymology of one-carbon metabolism in methanogenic pathways. *FEMS Microbiol Rev*, *23*(1), 13-38.

Fiedorek, S. C., Pumphrey, C. L., & Casteel, H. B. (1990). Breath methane production in children with constipation and encopresis. *J Pediatr Gastroenterol Nutr*, *10*(4), 473-477.

Finch, L. K., Ling, R., Napthine, S., Olspert, A., Michiels, T., Lardinois, C., Bell, S., Loughran, G., Brierley, I., & Firth, A. E. (2015). Characterization of Ribosomal Frameshifting in Theiler's Murine Encephalomyelitis Virus. *J Virol*, *89*(16), 8580-8589. https://doi.org/10.1128/jvi.01043-15

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Tilman, D., & Zaks, D. P. (2011). Solutions for a cultivated planet. *Nature*, *478*(7369), 337-342. https://doi.org/10.1038/nature10452

Franzosa, E. A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H. J., Reinker, S., Vatanen, T., Hall, A. B., Mallick, H., McIver, L. J., Sauk, J. S., Wilson, R. G., Stevens, B. W., Scott, J. M., Pierce, K., Deik, A. A., Bullock, K., Imhann, F., Porter, J. A., Zhernakova, A., Fu, J., Weersma, R. K., Wijmenga, C., Clish, C. B., Vlamakis, H., Huttenhower, C., & Xavier, R. J. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol*, *4*(2), 293-305. https://doi.org/10.1038/s41564-018-0306-4

Fricke, W. F., Seedorf, H., Henne, A., Kruer, M., Liesegang, H., Hedderich, R., Gottschalk, G., & Thauer, R. K. (2006). The genome sequence of Methanosphaera stadtmanae reveals why this human intestinal archaeon is restricted to methanol and H2 for methane formation and ATP synthesis. *J Bacteriol*, *188*(2), 642-658. https://doi.org/10.1128/jb.188.2.642-658.2006

Fukushima, T., Usami, R., & Kamekura, M. (2007). A traditional Japanese-style salt field is a niche for haloarchaeal strains that can survive in 0.5% salt solution. *Saline Systems*, *3*, 2. https://doi.org/10.1186/1746-1448-3-2

Furnari, M., Savarino, E., Bruzzone, L., Moscatelli, A., Gemignani, L., Giannini, E. G., Zentilin, P., Dulbecco, P., & Savarino, V. (2012). Reassessment of the role of methane production

between irritable bowel syndrome and functional constipation. *J Gastrointestin Liver Dis*, *21*(2), 157-163. https://www.jgld.ro/jgld/index.php/jgld/article/download/2012.2.10/727

Gaci, N., Borrel, G., Tottey, W., O'Toole, P. W., & Brugere, J. F. (2014). Archaea and the human gut: New beginning of an old story. *World Journal of Gastroenterology*, *20*(43), 16062-16078. https://doi.org/10.3748/wjg.v20.i43.16062

Gagen, E. J., Wang, J., Padmanabha, J., Liu, J., de Carvalho, I. P., Liu, J., Webb, R. I., Al Jassim, R., Morrison, M., Denman, S. E., & McSweeney, C. S. (2014). Investigation of a new acetogen isolated from an enrichment of the tammar wallaby forestomach. *BMC Microbiol*, *14*, 314. https://doi.org/10.1186/s12866-014-0314-3

Garcia, J. L., Patel, B. K., & Ollivier, B. (2000). Taxonomic, phylogenetic, and ecological diversity of methanogenic Archaea. *Anaerobe*, *6*(4), 205-226. https://doi.org/10.1006/anae.2000.0345

Ghanbari, M., Klose, V., Crispie, F., & Cotter, P. D. (2019). The dynamics of the antibiotic resistome in the feces of freshly weaned pigs following therapeutic administration of oxytetracycline. *Sci Rep*, *9*(1), 4062. https://doi.org/10.1038/s41598-019-40496-8

Ghavami, S. B., Rostami, E., Sephay, A. A., Shahrokh, S., Balaii, H., Aghdaei, H. A., & Zali, M. R. (2018). Alterations of the human gut Methanobrevibacter smithii as a biomarker for inflammatory bowel diseases. *Microb Pathog*, *117*, 285-289. https://doi.org/10.1016/j.micpath.2018.01.029

Ghoshal, U., Shukla, R., Srivastava, D., & Ghoshal, U. C. (2016). Irritable Bowel Syndrome, Particularly the Constipation-Predominant Form, Involves an Increase in Methanobrevibacter smithii, Which Is Associated with Higher Methane Production. *Gut Liver*, *10*(6), 932-938. https://doi.org/10.5009/gnl15588

Ghoshal, U. C., Srivastava, D., & Misra, A. (2018). A randomized double-blind placebo-controlled trial showing rifaximin to improve constipation by reducing methane production and accelerating colon transit: A pilot study. *Indian J Gastroenterol*, *37*(5), 416-423. https://doi.org/10.1007/s12664-018-0901-6

Ghoshal, U. C., Srivastava, D., Verma, A., & Misra, A. (2011). Slow transit constipation associated with excess methane production and its improvement following rifaximin therapy: a case report. *J Neurogastroenterol Motil*, *17*(2), 185-188. https://doi.org/10.5056/jnm.2011.17.2.185

Gibson, K. M., Nguyen, B. N., Neumann, L. M., Miller, M., Buss, P., Daniels, S., Ahn, M. J., Crandall, K. A., & Pukazhenthi, B. (2019). Gut microbiome differences between wild and captive black rhinoceros – implications for rhino health. *Sci Rep*, *9*(1), 7570. https://doi.org/10.1038/s41598-019-43875-3

Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., & Knight, R. (2018). Current understanding of the human microbiome [Review Article]. *Nat Med*, *24*, 392. https://doi.org/10.1038/nm.4517

Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M., & Nelson, K. E. (2006). Metagenomic analysis of the human distal gut microbiome. *Science*, *312*(5778), 1355-1359. https://doi.org/10.1126/science.1124234

Gilmore, S. P., Henske, J. K., Sexton, J. A., Solomon, K. V., Seppala, S., Yoo, J. I., Huyett, L. M., Pressman, A., Cogan, J. Z., Kivenson, V., Peng, X., Tan, Y., Valentine, D. L., & O'Malley, M. A. (2017). Genomic analysis of methanogenic archaea reveals a shift towards energy conservation. *BMC Genomics*, *18*(1), 639. https://doi.org/10.1186/s12864-017-4036-4

Girardin, S. E., Boneca, I. G., Carneiro, L. A., Antignac, A., Jehanno, M., Viala, J., Tedin, K., Taha, M. K., Labigne, A., Zahringer, U., Coyle, A. J., DiStefano, P. S., Bertin, J., Sansonetti, P. J., & Philpott, D. J. (2003). Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan. *Science*, *300*(5625), 1584-1587. https://doi.org/10.1126/science.1084677

Girardin, S. E., Boneca, I. G., Viala, J., Chamaillard, M., Labigne, A., Thomas, G., Philpott, D. J., & Sansonetti, P. J. (2003). Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *J Biol Chem*, *278*(11), 8869-8872. https://doi.org/10.1074/jbc.C200651200

Glaeser, S. P., Silva, L. M. R., Prieto, R., Silva, M. A., Franco, A., Kämpfer, P., Hermosilla, C., Taubert, A., & Eisenberg, T. (2021). A Preliminary Comparison on Faecal Microbiomes of Free-Ranging Large Baleen (Balaenoptera musculus, B. physalus, B. borealis) and Toothed (Physeter macrocephalus) Whales. *Microbial Ecology*. https://doi.org/10.1007/s00248-021-01729-4

Glendinning, L., Genç, B., Wallace, R. J., & Watson, M. (2021). Metagenomic analysis of the cow, sheep, reindeer and red deer rumen. *Sci Rep*, *11*(1), 1990. https://doi.org/10.1038/s41598-021-81668-9

Goenrich, M., Thauer, R. K., Yurimoto, H., & Kato, N. (2005). Formaldehyde activating enzyme (Fae) and hexulose-6-phosphate synthase (Hps) in Methanosarcina barkeri: a possible function in ribose-5-phosphate biosynthesis. *Arch Microbiol*, *184*(1), 41-48. https://doi.org/10.1007/s00203-005-0008-1

Gong, G., Zhou, S., Luo, R., Gesang, Z., & Suolang, S. (2020). Metagenomic insights into the diversity of carbohydrate-degrading enzymes in the yak fecal microbial community. *BMC Microbiol*, *20*(1), 302. https://doi.org/10.1186/s12866-020-01993-3

Gorlas, A., Robert, C., Gimenez, G., Drancourt, M., & Raoult, D. (2012). Complete genome sequence of Methanomassiliicoccus luminyensis, the largest genome of a human-associated Archaea species. *J Bacteriol*, *194*(17), 4745. https://doi.org/10.1128/jb.00956-12

Graham, D. E., & White, R. H. (2002). Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics [10.1039/B103714P]. *Natural Product Reports*, *19*(2), 133-147. https://doi.org/10.1039/B103714P

Grine, G., Drouet, H., Fenollar, F., Bretelle, F., Raoult, D., & Drancourt, M. (2019). Detection of Methanobrevibacter smithii in vaginal samples collected from women diagnosed with bacterial vaginosis. *Eur J Clin Microbiol Infect Dis*, *38*(9), 1643-1649. https://doi.org/10.1007/s10096-019-03592-1

Grine, G., Lotte, R., Chirio, D., Chevalier, A., Raoult, D., Drancourt, M., & Ruimy, R. (2019). Co-culture of Methanobrevibacter smithii with enterobacteria during urinary infection. *EBioMedicine*, *43*, 333-337. https://doi.org/https://doi.org/10.1016/j.ebiom.2019.04.037

Grossi, G., Goglio, P., Vitali, A., & Williams, A. G. (2018). Livestock and climate change: impact of livestock on climate and mitigation strategies. *Animal Frontiers*, *9*(1), 69-76. https://doi.org/10.1093/af/vfy034

Gu, P., Patel, D., Lakhoo, K., Ko, J., Liu, X., Chang, B., Pan, D., Lentz, G., Sonesen, M., Estiandan, R., Lin, E., Pimentel, M., & Rezaie, A. (2020). Breath Test Gas Patterns in Inflammatory Bowel Disease with Concomitant Irritable Bowel Syndrome-Like Symptoms: A Controlled Large-Scale Database Linkage Analysis. *Dig Dis Sci*, *65*(8), 2388-2396. https://doi.org/10.1007/s10620-019-05967-y

Guitor, A. K., Raphenya, A. R., Klunk, J., Kuch, M., Alcock, B., Surette, M. G., McArthur, A. G., Poinar, H. N., & Wright, G. D. (2019). Capturing the Resistome: a Targeted Capture Method To Reveal Antibiotic Resistance Determinants in Metagenomes. *Antimicrob Agents Chemother*, *64*(1). https://doi.org/10.1128/aac.01324-19

Gulino, L. M., Ouwerkerk, D., Kang, A. Y., Maguire, A. J., Kienzle, M., & Klieve, A. V. (2013). Shedding light on the microbial community of the macropod foregut using 454-amplicon pyrosequencing. *Plos One*, *8*(4), e61463. https://doi.org/10.1371/journal.pone.0061463

Guo, W., Mishra, S., Wang, C., Zhang, H., Ning, R., Kong, F., Zeng, B., Zhao, J., & Li, Y. (2019). Comparative Study of Gut Microbiota in Wild and Captive Giant Pandas (Ailuropoda melanoleuca). *Genes (Basel)*, *10*(10). https://doi.org/10.3390/genes10100827

Haines, A., Metz, G., Dilawari, J., Blendis, L., & Wiggins, H. (1977). Breath-methane in patients with cancer of the large bowel. *Lancet*, *2*(8036), 481-483.

Hakki, Z., Thompson, A. J., Bellmaine, S., Speciale, G., Davies, G. J., & Williams, S. J. (2015). Structural and kinetic dissection of the endo-α-1,2-mannanase activity of bacterial GH99

glycoside hydrolases from Bacteroides spp. *Chemistry*, *21*(5), 1966-1977. https://doi.org/10.1002/chem.201405539

Hall, A. B., Yassour, M., Sauk, J., Garner, A., Jiang, X., Arthur, T., Lagoudas, G. K., Vatanen, T., Fornelos, N., Wilson, R., Bertha, M., Cohen, M., Garber, J., Khalili, H., Gevers, D., Ananthakrishnan, A. N., Kugathasan, S., Lander, E. S., Blainey, P., Vlamakis, H., Xavier, R. J., & Huttenhower, C. (2017). A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. *Genome Med*, *9*(1), 103. https://doi.org/10.1186/s13073-017-0490-5

Hansen, E. E., Lozupone, C. A., Rey, F. E., Wu, M., Guruge, J. L., Narra, A., Goodfellow, J., Zaneveld, J. R., McDonald, D. T., Goodrich, J. A., Heath, A. C., Knight, R., & Gordon, J. I. (2011). Pan-genome of the dominant human gut-associated archaeon, Methanobrevibacter smithii, studied in twins. *Proc Natl Acad Sci U S A*, *108 Suppl 1*, 4599-4606. https://doi.org/10.1073/pnas.1000071108

Hassani, Y., Brégeon, F., Aboudharam, G., Drancourt, M., & Grine, G. (2020). Detection of Methanobrevobacter smithii and Methanobrevibacter oralis in Lower Respiratory Tract Microbiota. *Microorganisms*, *8*(12). https://doi.org/10.3390/microorganisms8121866

Hatzenpichler, R. (2012). Diversity, physiology, and niche differentiation of ammonia-oxidizing archaea. *Appl Environ Microbiol*, *78*(21), 7501-7510. https://doi.org/10.1128/aem.01960-12

He, F., Liu, D., Zhang, L., Zhai, J., Ma, Y., Xu, Y., Jiang, G., Rong, K., & Ma, J. (2018). Metagenomic analysis of captive Amur tiger faecal microbiome. *BMC Vet Res*, *14*(1), 379. https://doi.org/10.1186/s12917-018-1696-5

Heidarian, F., Alebouyeh, M., Shahrokh, S., Balaii, H., & Zali, M. R. (2019). Altered fecal bacterial composition correlates with disease activity in inflammatory bowel disease and the extent of IL8 induction. *Current Research in Translational Medicine*, *67*(2), 41-50. https://doi.org/https://doi.org/10.1016/j.retram.2019.01.002

Heinken, A., Ravcheev, D. A., Baldini, F., Heirendt, L., Fleming, R. M. T., & Thiele, I. (2019). Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease. *Microbiome*, *7*(1), 75. https://doi.org/10.1186/s40168-019-0689-3

Hoedt, E. C., Ó Cuív, P., Evans, P. N., Smith, W. J. M., McSweeney, C. S., Denman, S. E., & Morrison, M. (2016). Differences down-under: alcohol-fueled methanogenesis by archaea present in Australian macropodids. *The ISME Journal*, *10*(10), 2376-2388. https://doi.org/10.1038/ismej.2016.41

Hoedt, E. C., Parks, D. H., Volmer, J. G., Rosewarne, C. P., Denman, S. E., McSweeney, C. S., Muir, J. G., Gibson, P. R., Cuiv, P. O., Hugenholtz, P., Tyson, G. W., & Morrison, M.

(2018). Culture- and metagenomics-enabled analyses of the Methanosphaera genus reveals their monophyletic origin and differentiation according to genome size. *Isme j*, *12*(12), 2942-2953. https://doi.org/10.1038/s41396-018-0225-7

Hoff, G., Bjorneklett, A., Moen, I. E., & Jenssen, E. (1986). Epidemiology of polyps in the rectum and sigmoid colon. Evaluation of breath methane and predisposition for colorectal neoplasia. *Scand J Gastroenterol*, *21*(2), 193-198. https://www.tandfonline.com/doi/pdf/10.3109/00365528609034646?needAccess=true

Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G. D., Lewis, J. D., & Bushman, F. D. (2013). Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. *Plos One*, *8*(6), e66019. https://doi.org/10.1371/journal.pone.0066019

Hong, P.-Y., Wheeler, E., Cann, I. K. O., & Mackie, R. I. (2011). Phylogenetic analysis of the fecal microbial community in herbivorous land and marine iguanas of the Galápagos Islands using 16S rRNA-based pyrosequencing. *The ISME Journal*, *5*(9), 1461-1470. https://doi.org/10.1038/ismej.2011.33

Horz, H. P., Seyfarth, I., & Conrads, G. (2012). McrA and 16S rRNA gene analysis suggests a novel lineage of Archaea phylogenetically affiliated with Thermoplasmatales in human subgingival plaque. *Anaerobe*, *18*(3), 373-377. https://doi.org/10.1016/j.anaerobe.2012.04.006

Hou, Q., Kwok, L.-Y., Zheng, Y., Wang, L., Guo, Z., Zhang, J., Huang, W., Wang, Y., Leng, L., Li, H., & Zhang, H. (2016). Differential fecal microbiota are retained in broiler chicken lines divergently selected for fatness traits. *Sci Rep*, *6*, 37376-37376. https://doi.org/10.1038/srep37376

Hoyles, L., Jiménez-Pranteda, M. L., Chilloux, J., Brial, F., Myridakis, A., Aranias, T., Magnan, C., Gibson, G. R., Sanderson, J. D., Nicholson, J. K., Gauguier, D., McCartney, A. L., & Dumas, M.-E. (2018). Metabolic retroconversion of trimethylamine N-oxide and the gut microbiota. *Microbiome*, *6*(1), 73. https://doi.org/10.1186/s40168-018-0461-0

Hu, P.-L., Yuan, Y.-H., Yue, T.-L., & Guo, C.-F. (2018). Bile acid patterns in commercially available oxgall powders used for the evaluation of the bile tolerance ability of potential probiotics. *Plos One*, *13*(3), e0192964. https://doi.org/10.1371/journal.pone.0192964

Huang, X. D., Tan, H. Y., Long, R., Liang, J. B., & Wright, A. D. (2012). Comparison of methanogen diversity of yak (Bos grunniens) and cattle (Bos taurus) from the Qinghai-Tibetan plateau, China. *BMC Microbiol*, *12*, 237. https://doi.org/10.1186/1471-2180-12-237

Huebner, K. L., Martin, J. N., Weissend, C. J., Holzer, K. L., Parker, J. K., Lakin, S. M., Doster, E., Weinroth, M. D., Abdo, Z., Woerner, D. R., Metcalf, J. L., Geornaras, I., Bryant, T. C., Morley, P. S., & Belk, K. E. (2019). Effects of a Saccharomyces cerevisiae fermentation

product on liver abscesses, fecal microbiome, and resistome in feedlot cattle raised without antibiotics. *Sci Rep*, *9*(1), 2559. https://doi.org/10.1038/s41598-019-39181-7

Human Microbiome Project, C. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), 207-214. https://doi.org/10.1038/nature11234

Hungate, R. E., & Macy, J. (1973). The Roll-Tube Method for Cultivation of Strict Anaerobes. *Bulletins from the Ecological Research Committee*(17), 123-126. www.jstor.org/stable/20111550

Huntemann, M., Ivanova, N. N., Mavromatis, K., Tripp, H. J., Paez-Espino, D., Palaniappan, K., Szeto, E., Pillay, M., Chen, I. M., Pati, A., Nielsen, T., Markowitz, V. M., & Kyrpides, N. C. (2015). The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Stand Genomic Sci*, *10*, 86. https://doi.org/10.1186/s40793-015-0077-y

Huppe, C. A., Blais Lecours, P., Lechasseur, A., Gendron, D. R., Lemay, A. M., Bissonnette, E. Y., Blanchet, M. R., Duchaine, C., Morissette, M. C., Rosen, H., & Marsolais, D. (2018). A sphingosine-1-phosphate receptor 1 agonist inhibits tertiary lymphoid tissue reactivation and hypersensitivity in the lung. *Mucosal Immunol*, *11*(1), 112-119. https://doi.org/10.1038/mi.2017.37

Huseby, D. L., Brandis, G., Praski Alzrigat, L., & Hughes, D. (2020). Antibiotic resistance by high-level intrinsic suppression of a frameshift mutation in an essential gene. *Proc Natl Acad Sci U S A*, *117*(6), 3185-3191. https://doi.org/10.1073/pnas.1919390117

Huynh, H. T. T., Pignoly, M., Drancourt, M., & Aboudharam, G. (2017). A new methanogen "Methanobrevibacter massiliense" isolated in a case of severe periodontitis. *BMC Res Notes*, *10*(1), 657. https://doi.org/10.1186/s13104-017-2980-3

Huynh, H. T. T., Pignoly, M., Nkamga, V. D., Drancourt, M., & Aboudharam, G. (2015). The Repertoire of Archaea Cultivated from Severe Periodontitis. *Plos One*, *10*(4), e0121565. https://doi.org/10.1371/journal.pone.0121565

Hwang, L., Low, K., Khoshini, R., Melmed, G., Sahakian, A., Makhani, M., Pokkunuri, V., & Pimentel, M. (2010). Evaluating breath methane as a diagnostic test for constipation-predominant IBS. *Dig Dis Sci*, *55*(2), 398-403. https://doi.org/10.1007/s10620-009-0778-4

Iino, T., Tamaki, H., Tamazawa, S., Ueno, Y., Ohkuma, M., Suzuki, K., Igarashi, Y., & Haruta, S. (2013). Candidatus Methanogranum caenicola: a novel methanogen from the anaerobic digested sludge, and proposal of Methanomassiliicoccaceae fam. nov. and Methanomassiliicoccales ord. nov., for a methanogenic lineage of the class Thermoplasmata. *Microbes Environ*, *28*(2), 244-250. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4070666/pdf/28_244.pdf

Ilmberger, N., Güllert, S., Dannenberg, J., Rabausch, U., Torres, J., Wemheuer, B., Alawi, M., Poehlein, A., Chow, J., Turaev, D., Rattei, T., Schmeisser, C., Salomon, J., Olsen, P. B., Daniel, R., Grundhoff, A., Borchert, M. S., & Streit, W. R. (2014). A comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed Asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes. *Plos One*, *9*(9), e106707. https://doi.org/10.1371/journal.pone.0106707

Jahng, J., Jung, I. S., Choi, E. J., Conklin, J. L., & Park, H. (2012). The effects of methane and hydrogen gases produced by enteric bacteria on ileal motility and colonic transit time. *Neurogastroenterol Motil*, *24*(2), 185-190, e192. https://doi.org/10.1111/j.1365-2982.2011.01819.x

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*, *9*(1), 5114. https://doi.org/10.1038/s41467-018-07641-9

Jangi, S., Gandhi, R., Cox, L. M., Li, N., von Glehn, F., Yan, R., Patel, B., Mazzola, M. A., Liu, S., Glanz, B. L., Cook, S., Tankou, S., Stuart, F., Melo, K., Nejad, P., Smith, K., Topçuolu, B. D., Holden, J., Kivisäkk, P., Chitnis, T., De Jager, P. L., Quintana, F. J., Gerber, G. K., Bry, L., & Weiner, H. L. (2016). Alterations of the human gut microbiome in multiple sclerosis. *Nat Commun*, *7*(1), 12015. https://doi.org/10.1038/ncomms12015

Jaworska, K., Hering, D., Mosieniak, G., Bielak-Zmijewska, A., Pilz, M., Konwerski, M., Gasecka, A., Kapłon-Cieślicka, A., Filipiak, K., Sikora, E., Hołyst, R., & Ufnal, M. (2019). TMA, A Forgotten Uremic Toxin, but Not TMAO, Is Involved in Cardiovascular Pathology. *Toxins (Basel)*, *11*(9). https://doi.org/10.3390/toxins11090490

Jensen, N. S., & Canale-Parola, E. (1986). Bacteroides pectinophilus sp. nov. and Bacteroides galacturonicus sp. nov.: two pectinolytic bacteria from the human intestinal tract. *Appl Environ Microbiol*, *52*(4), 880-887. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC239131/pdf/aem00133-0286.pdf

Jhangi, S., Gandhi, R., Glanz, B., Cook, S., Nejad, P., Ward, D., Li, N., Gerber, G., Bry, L., & Weiner, H. (2014). Increased Archaea Species and Changes with Therapy in Gut Microbiome of Multiple Sclerosis Subjects (S24.001). *Neurology*, *82*(10 Supplement), S24.001.

Jie, Z., Xia, H., Zhong, S.-L., Feng, Q., Li, S., Liang, S., Zhong, H., Liu, Z., Gao, Y., Zhao, H., Zhang, D., Su, Z., Fang, Z., Lan, Z., Li, J., Xiao, L., Li, J., Li, R., Li, X., Li, F., Ren, H., Huang, Y., Peng, Y., Li, G., Wen, B., Dong, B., Chen, J.-Y., Geng, Q.-S., Zhang, Z.-W., Yang, H., Wang, J., Wang, J., Zhang, X., Madsen, L., Brix, S., Ning, G., Xu, X., Liu, X., Hou, Y., Jia, H., He, K., & Kristiansen, K. (2017). The gut microbiome in atherosclerotic

cardiovascular disease. *Nat Commun*, *8*(1), 845. https://doi.org/10.1038/s41467-017-00900-1

Joblin, K. N., Naylor, G. E., & Williams, A. G. (1990). Effect of Methanobrevibacter-smithii on Xylanolytic Activity of Anaerobic Ruminal Fungi. *Applied and Environmental Microbiology*, *56*(8), 2287-2295.

Johnson, C., England, A., Munro-Ehrlich, M., Colman, D. R., DuBois, J. L., & Boyd, E. S. (2021). Pathways of Iron and Sulfur Acquisition, Cofactor Assembly, Destination, and Storage in Diverse Archaeal Methanogens and Alkanotrophs. *J Bacteriol*, *203*(17), e0011721. https://doi.org/10.1128/jb.00117-21

Johnson, K. A., & Johnson, D. E. (1995). Methane emissions from cattle. *Journal of Animal Science*, *73*(8), 2483-2492. https://doi.org/10.2527/1995.7382483x

Jones, B. V., Begley, M., Hill, C., Gahan, C. G., & Marchesi, J. R. (2008). Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proc Natl Acad Sci U S A*, *105*(36), 13580-13585. https://doi.org/10.1073/pnas.0804437105

Joulian, C., Ollivier, B., Patel, B. K. C., & Roger, P. A. (1998). Phenotypic and phylogenetic characterization of dominant culturable methanogens isolated from ricefield soils. *FEMS Microbiology Ecology*, *25*(2), 135-145. https://doi.org/https://doi.org/10.1016/S0168-6496(97)00090-1

Joyce, S. A., & Gahan, C. G. (2017). Disease-Associated Changes in Bile Acid Profiles and Links to Altered Gut Microbiota. *Dig Dis*, *35*(3), 169-177. https://doi.org/10.1159/000450907

Kamil, R. Z., Murdiati, A., Juffrie, M., Nakayama, J., & Rahayu, E. S. (2021). Gut Microbiota and Short-Chain Fatty Acid Profile between Normal and Moderate Malnutrition Children in Yogyakarta, Indonesia. *Microorganisms*, *9*(1). https://doi.org/10.3390/microorganisms9010127

Kamke, J., Kittelmann, S., Soni, P., Li, Y., Tavendale, M., Ganesh, S., Janssen, P. H., Shi, W., Froula, J., Rubin, E. M., & Attwood, G. T. (2016). Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea-enriched microbiome characterised by lactic acid formation and utilisation. *Microbiome*, *4*(1), 56. https://doi.org/10.1186/s40168-016-0201-2

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*, *44*(D1), D457-D462. https://doi.org/10.1093/nar/gkv1070

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol*, *428*(4), 726-731. https://doi.org/10.1016/j.jmb.2015.11.006

Kang, D. D., Froula, J., Egan, R., & Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, *3*, e1165-e1165. https://doi.org/10.7717/peerj.1165

Karlin, D. A., Jones, R. D., Stroehlein, J. R., Mastromarino, A. J., & Potter, G. D. (1982). Breath methane excretion in patients with unresected colorectal cancer. *J Natl Cancer Inst*, *69*(3), 573-576.

Kashtan, H., Rabau, M., Peled, Y., Milstein, A., & Wiznitzer, T. (1989). Methane production in patients with colorectal carcinoma. *Isr J Med Sci*, *25*(11), 614-616.

Kataoka, H., Yasuda, M., Iyori, M., Kiura, K., Narita, M., Nakata, T., & Shibata, K. (2006). Roles of N-linked glycans in the recognition of microbial lipopeptides and lipoproteins by TLR2. *Cell Microbiol*, *8*(7), 1199-1209. https://doi.org/10.1111/j.1462-5822.2006.00702.x

Katayama, N., Takeya, M., & Osanai, T. (2019). Biochemical characterisation of fumarase C from a unicellular cyanobacterium demonstrating its substrate affinity, altered by an amino acid substitution. *Sci Rep*, *9*(1), 10629. https://doi.org/10.1038/s41598-019-47025-7

Kc, K. B., Dias, G. M., Veeramani, A., Swanton, C. J., Fraser, D., Steinke, D., Lee, E., Wittman, H., Farber, J. M., Dunfield, K., McCann, K., Anand, M., Campbell, M., Rooney, N., Raine, N. E., Acker, R. V., Hanner, R., Pascoal, S., Sharif, S., Benton, T. G., & Fraser, E. D. G. (2018). When too much isn't enough: Does current food production meet global nutritional needs? *Plos One*, *13*(10), e0205683. https://doi.org/10.1371/journal.pone.0205683

Kelly, W. J., Li, D., Lambie, S. C., Jeyanathan, J., Cox, F., Li, Y., Attwood, G. T., Altermann, E., & Leahy, S. C. (2016). Complete Genome Sequence of Methanogenic Archaeon ISO4-G1, a Member of the Methanomassiliicoccales, Isolated from a Sheep Rumen. *Genome Announc*, *4*(2). https://doi.org/10.1128/genomeA.00221-16

Ketteler, R. (2012). On programmed ribosomal frameshifting: the alternative proteomes [10.3389/fgene.2012.00242]. *Front Genet*, *3*, 242. https://www.frontiersin.org/article/10.3389/fgene.2012.00242

Khan, A. R., & Awan, F. R. (2014). Metals in the pathogenesis of type 2 diabetes. *J Diabetes Metab Disord*, *13*(1), 16. https://doi.org/10.1186/2251-6581-13-16

Khelaifia, S., Caputo, A., Djossou, F., & Raoult, D. (2017). Draft genome sequence of a human-associated isolate of Haloferax alexandrinus strain Arc-hr, an extremely halophilic archaea. *New Microbes New Infect*, *15*, 44-45. https://doi.org/10.1016/j.nmni.2016.11.012

Khelaifia, S., Garibal, M., Robert, C., Raoult, D., & Drancourt, M. (2014). Draft Genome Sequence of a Human-Associated Isolate of Methanobrevibacter arboriphilicus, the Lowest-G+C-Content Archaeon. *Genome Announc*, *2*(1). https://doi.org/10.1128/genomeA.01181-13

Khelaifia, S., & Raoult, D. (2016). Haloferax massiliensis sp. nov., the first human-associated halophilic archaea. *New Microbes New Infect*, *12*, 96-98. https://doi.org/10.1016/j.nmni.2016.05.007

Kim, B. C., & Jeong, H. (2018). Complete Genome Sequence of Methanobrevibacter smithii Strain KB11, Isolated from a Korean Fecal Sample. *Genome Announc*, *6*(7). https://doi.org/10.1128/genomeA.00038-18

Kim, G., Deepinder, F., Morales, W., Hwang, L., Weitsman, S., Chang, C., Gunsalus, R., & Pimentel, M. (2012). Methanobrevibacter smithii is the predominant methanogen in patients with constipation-predominant IBS and methane on breath. *Dig Dis Sci*, *57*(12), 3213-3218. https://doi.org/10.1007/s10620-012-2197-1

Kim, J. Y., Whon, T. W., Lim, M. Y., Kim, Y. B., Kim, N., Kwon, M.-S., Kim, J., Lee, S. H., Choi, H.-J., Nam, I.-H., Chung, W.-H., Kim, J.-H., Bae, J.-W., Roh, S. W., & Nam, Y.-D. (2020). The human gut archaeome: identification of diverse haloarchaea in Korean subjects. *Microbiome*, *8*(1), 114. https://doi.org/10.1186/s40168-020-00894-x

Kisiela, M., Skarka, A., Ebert, B., & Maser, E. (2012). Hydroxysteroid dehydrogenases (HSDs) in bacteria: a bioinformatic perspective. *J Steroid Biochem Mol Biol*, *129*(1-2), 31-46. https://doi.org/10.1016/j.jsbmb.2011.08.002

Kitamura, K., Fujita, T., Akada, S., & Tonouchi, A. (2011). Methanobacterium kanagiense sp. nov., a hydrogenotrophic methanogen, isolated from rice-field soil. *International Journal of Systematic and Evolutionary Microbiology*, *61*(6), 1246-1252. https://doi.org/https://doi.org/10.1099/ijs.0.026013-0

Klieve, A. V., Ouwerkerk, D., & Maguire, A. J. (2012). Archaea in the foregut of macropod marsupials: PCR and amplicon sequence-based observations. *J Appl Microbiol*, *113*(5), 1065-1075. https://doi.org/10.1111/j.1365-2672.2012.05428.x

Kobayashi, T., Kimura, B., & Fujii, T. (2000). Haloanaerobium fermentans sp. nov., a strictly anaerobic, fermentative halophile isolated from fermented puffer fish ovaries. *Int J Syst Evol Microbiol*, *50 Pt 4*, 1621-1627. https://doi.org/10.1099/00207713-50-4-1621

Koeth, R. A., Wang, Z., Levison, B. S., Buffa, J. A., Org, E., Sheehy, B. T., Britt, E. B., Fu, X., Wu, Y., Li, L., Smith, J. D., DiDonato, J. A., Chen, J., Li, H., Wu, G. D., Lewis, J. D., Warrier, M., Brown, J. M., Krauss, R. M., Tang, W. H., Bushman, F. D., Lusis, A. J., & Hazen, S. L. (2013). Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med*, *19*(5), 576-585. https://doi.org/10.1038/nm.3145

Kohl, K. D., Miller, A. W., Marvin, J. E., Mackie, R., & Dearing, M. D. (2014). Herbivorous rodents (Neotoma spp.) harbour abundant and active foregut microbiota. *Environ Microbiol*, *16*(9), 2869-2878. https://doi.org/10.1111/1462-2920.12376

Konstantinidis, K. T., & Tiedje, J. M. (2005). Towards a genome-based taxonomy for prokaryotes. *J Bacteriol*, *187*(18), 6258-6264. https://doi.org/10.1128/jb.187.18.6258-6264.2005

Koskinen, K., Pausan, M. R., Perras, A. K., Beck, M., Bang, C., Mora, M., Schilhabel, A., Schmitz, R., & Moissl-Eichinger, C. (2017). First Insights into the Diverse Human Archaeome: Specific Detection of Archaea in the Gastrointestinal Tract, Lung, and Nose and on Skin [10.1128/mBio.00824-17]. *mBio*, *8*(6). http://mbio.asm.org/content/8/6/e00824-17.abstract

Koskinen, K., Reichert, J. L., Hoier, S., Schachenreiter, J., Duller, S., Moissl-Eichinger, C., & Schöpf, V. (2018). The nasal microbiome mirrors and potentially shapes olfactory function. *Sci Rep*, *8*(1), 1296. https://doi.org/10.1038/s41598-018-19438-3

Kulik, E. M., Sandmeier, H., Hinni, K., & Meyer, J. (2001). Identification of archaeal rDNA from subgingival dental plaque by PCR amplification and sequence analysis. *FEMS Microbiol Lett*, *196*(2), 129-133.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*, *35*(6), 1547-1549. https://doi.org/10.1093/molbev/msy096

Kurth, J. M., Op den Camp, H. J. M., & Welte, C. U. (2020). Several ways one goal—methanogenesis from unconventional substrates. *Appl Microbiol Biotechnol*, *104*(16), 6839-6854. https://doi.org/10.1007/s00253-020-10724-7

Kusar, D., & Avgustin, G. (2010). Molecular profiling and identification of methanogenic archaeal species from rabbit caecum. *FEMS Microbiol Ecol*, *74*(3), 623-630. https://doi.org/10.1111/j.1574-6941.2010.00980.x

Lassey, K. R. (2007). Livestock methane emission: from the individual grazing animal through national inventories to the global methane cycle. *Agricultural and forest meteorology*, *142*(2-4), 120-132.

Lau, W. L., Savoj, J., Nakata, M. B., & Vaziri, N. D. (2018). Altered microbiome in chronic kidney disease: systemic effects of gut-derived uremic toxins. *Clin Sci (Lond)*, *132*(5), 509-522. https://doi.org/10.1042/cs20171107

Leahy, S. C., Kelly, W. J., Altermann, E., Ronimus, R. S., Yeoman, C. J., Pacheco, D. M., Li, D., Kong, Z., McTavish, S., Sang, C., Lambie, S. C., Janssen, P. H., Dey, D., & Attwood, G. T. (2010). The genome sequence of the rumen methanogen Methanobrevibacter ruminantium reveals new possibilities for controlling ruminant methane emissions. *Plos One*, *5*(1), e8926. https://doi.org/10.1371/journal.pone.0008926

Lee, J. M., Lee, K. M., Chung, Y. Y., Lee, Y. W., Kim, D. B., Sung, H. J., Chung, W. C., & Paik, C. N. (2015). Clinical significance of the glucose breath test in patients with inflammatory bowel disease. *J Gastroenterol Hepatol*, *30*(6), 990-994. https://doi.org/10.1111/jgh.12908

Lee, K. M., Paik, C. N., Chung, W. C., Yang, J. M., & Choi, M. G. (2013). Breath methane positivity is more common and higher in patients with objectively proven delayed transit constipation. *Eur J Gastroenterol Hepatol*, *25*(6), 726-732. https://doi.org/10.1097/MEG.0b013e32835eb916

Lee, S., Cantarel, B., Henrissat, B., Gevers, D., Birren, B. W., Huttenhower, C., & Ko, G. (2014). Gene-targeted metagenomic analysis of glucan-branching enzyme gene profiles among human and animal fecal microbiota. *Isme j*, *8*(3), 493-503. https://doi.org/10.1038/ismej.2013.167

Leipold, M. D., Vinogradov, E., & Whitfield, C. (2007). Glycosyltransferases involved in biosynthesis of the outer core region of Escherichia coli lipopolysaccharides exhibit broader substrate specificities than is predicted from lipopolysaccharide structures. *J Biol Chem*, *282*(37), 26786-26792. https://doi.org/10.1074/jbc.M704131200

Li, C., Tan, X., Bai, J., Xu, Q., Liu, S., Guo, W., Yu, C., Fan, G., Lu, Y., Zhang, H., Yang, H., Chen, J., & Liu, X. (2019). A survey of the sperm whale (Physeter catodon) commensal microbiome. *PeerJ*, *7*, e7257. https://doi.org/10.7717/peerj.7257

Li, C. L., Liu, D. L., Jiang, Y. T., Zhou, Y. B., Zhang, M. Z., Jiang, W., Liu, B., & Liang, J. P. (2009). Prevalence and molecular diversity of Archaea in subgingival pockets of periodontitis patients. *Oral Microbiol Immunol*, *24*(4), 343-346. https://doi.org/10.1111/j.1399-302X.2009.00514.x

Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674-1676. https://doi.org/10.1093/bioinformatics/btv033

Li, F., Chen, C., Wei, W., Wang, Z., Dai, J., Hao, L., Song, L., Zhang, X., Zeng, L., Du, H., Tang, H., Liu, N., Yang, H., Wang, J., Madsen, L., Brix, S., Kristiansen, K., Xu, X., Li, J., Wu, R., & Jia, H. (2018). The metagenome of the female upper reproductive tract. *Gigascience*, *7*(10). https://doi.org/10.1093/gigascience/giy107

Lim, M. Y., You, H. J., Yoon, H. S., Kwon, B., Lee, J. Y., Lee, S., Song, Y. M., Lee, K., Sung, J., & Ko, G. (2017). The effect of heritability and host genetics on the gut microbiota and metabolic syndrome. *Gut*, *66*(6), 1031-1038. https://doi.org/10.1136/gutjnl-2015-311326

Lim, S.-K., Kim, D., Moon, D.-C., Cho, Y., & Rho, M. (2020). Antibiotic resistomes discovered in the gut microbiomes of Korean swine and cattle. *Gigascience*, *9*(5). https://doi.org/10.1093/gigascience/giaa043

Lima, J., Auffret, M. D., Stewart, R. D., Dewhurst, R. J., Duthie, C. A., Snelling, T. J., Walker, A. W., Freeman, T. C., Watson, M., & Roehe, R. (2019). Identification of Rumen Microbial Genes Involved in Pathways Linked to Appetite, Growth, and Feed Conversion Efficiency in Cattle. *Front Genet*, *10*, 701. https://doi.org/10.3389/fgene.2019.00701

Lin, C., & Miller, T. L. (1998). Phylogenetic analysis of Methanobrevibacter isolated from feces of humans and other animals. *Arch Microbiol*, *169*(5), 397-403. https://doi.org/10.1007/s002030050589

Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., Andrews, E., Ajami, N. J., Bonham, K. S., Brislawn, C. J., Casero, D., Courtney, H., Gonzalez, A., Graeber, T. G., Hall, A. B., Lake, K., Landers, C. J., Mallick, H., Plichta, D. R., Prasad, M., Rahnavard, G., Sauk, J., Shungin, D., Vázquez-Baeza, Y., White, R. A., 3rd, Braun, J., Denson, L. A., Jansson, J. K., Knight, R., Kugathasan, S., McGovern, D. P. B., Petrosino, J. F., Stappenbeck, T. S., Winter, H. S., Clish, C. B., Franzosa, E. A., Vlamakis, H., Xavier, R. J., & Huttenhower, C. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*, *569*(7758), 655-662. https://doi.org/10.1038/s41586-019-1237-9

Lo Sasso, G., Khachatryan, L., Kondylis, A., Battey, J. N. D., Sierro, N., Danilova, N. A., Grigoryeva, T. V., Markelova, M. I., Khusnutdinova, D. R., Laikov, A. V., Salafutdinov, I. I., Romanova, Y. D., Siniagina, M. N., Vasiliev, I. Y., Boulygina, E. A., Solovyeva, V. V., Garanina, E. E., Kitaeva, K. V., Ivanov, K. Y., Chulpanova, D. S., Kletenkov, K. S., Valeeva, A. R., Odintsova, A. K., Ardatskaya, M. D., Abdulkhakov, R. A., Ivanov, N. V., Peitsch, M. C., Hoeng, J., & Abdulkhakov, S. R. (2020). Inflammatory Bowel Disease–Associated Changes in the Gut: Focus on Kazan Patients. *Inflamm Bowel Dis*, *27*(3), 418-433. https://doi.org/10.1093/ibd/izaa188

Lodola, A., Branduardi, D., De Vivo, M., Capoferri, L., Mor, M., Piomelli, D., & Cavalli, A. (2012). A Catalytic Mechanism for Cysteine N-Terminal Nucleophile Hydrolases, as Revealed by Free Energy Simulations. *Plos One*, *7*(2), e32397. https://doi.org/10.1371/journal.pone.0032397

Lokmer, A., Cian, A., Froment, A., Gantois, N., Viscogliosi, E., Chabé, M., & Ségurel, L. (2019). Use of shotgun metagenomics for the identification of protozoa in the gut microbiota of healthy individuals from worldwide populations with various industrialization levels. *Plos One*, *14*(2), e0211139. https://doi.org/10.1371/journal.pone.0211139

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*, *42*(Database issue), D490-495. https://doi.org/10.1093/nar/gkt1178

Luo, Y., Wright, A. D. G., Li, Y., Li, H., Yang, Q., Luo, L., & Yang, M. (2013). Diversity of methanogens in the hindgut of captive white rhinoceroses, Ceratotherium simum. *BMC Microbiol*, *13*, 207. https://doi.org/10.1186/1471-2180-13-207

Lurie-Weinberger, M. N., & Gophna, U. (2015). Archaea in and on the Human Body: Health Implications and Future Directions. *PLoS Pathog*, *11*(6), e1004833. https://doi.org/10.1371/journal.ppat.1004833

Lwin, K. O., & Matsui, H. (2014). Comparative analysis of the methanogen diversity in horse and pony by using mcrA gene and archaeal 16s rRNA gene clone libraries. *Archaea*, *2014*, 483574. https://doi.org/10.1155/2014/483574

Mack, I., Cuntz, U., Grämer, C., Niedermaier, S., Pohl, C., Schwiertz, A., Zimmermann, K., Zipfel, S., Enck, P., & Penders, J. (2016). Weight gain in anorexia nervosa does not ameliorate the faecal microbiota, branched chain fatty acid profiles and gastrointestinal complaints. *Sci Rep*, *6*(1), 26752. https://doi.org/10.1038/srep26752

Mackie, R., Aminov, R., Gaskins, H., & White, B. (2000). Molecular microbial ecology in gut ecosystems. 8th International Symposium on Microbial Ecology, Halifax, Canada (1998, August 9–14),

Maczulak, A. E., Wolin, M. J., & Miller, T. L. (1989). Increase in colonic methanogens and total anaerobes in aging rats. *Appl Environ Microbiol*, *55*(10), 2468-2473.

Madeira, F., Madhusoodanan, N., Lee, J., Tivey, A. R. N., & Lopez, R. (2019). Using EMBL-EBI Services via Web Interface and Programmatically via Web Services. *Current Protocols in Bioinformatics*, *66*(1), e74. https://doi.org/10.1002/cpbi.74

Maita, N., Nyirenda, J., Igura, M., Kamishikiryo, J., & Kohda, D. (2010). Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases. *J Biol Chem*, *285*(7), 4941-4950. https://doi.org/10.1074/jbc.M109.081752

Majewski, M., & McCallum, R. W. (2007). Results of small intestinal bacterial overgrowth testing in irritable bowel syndrome patients: clinical profiles and effects of antibiotic trial. *Adv Med Sci*, *52*, 139-142.

Maldonado-Valderrama, J., Wilde, P., Macierzanka, A., & Mackie, A. (2011). The role of bile salts in digestion. *Adv Colloid Interface Sci*, *165*(1), 36-46. https://doi.org/10.1016/j.cis.2010.12.002

Mand, T. D., Kulkarni, G., Metcalf, W. W., & Becker, A. (2018). Genetic, Biochemical, and Molecular Characterization of Methanosarcina barkeri Mutants Lacking Three Distinct Classes of Hydrogenase. *J Bacteriol*, *200*(20), e00342-00318. https://doi.org/doi:10.1128/JB.00342-18

Mao, S. Y., Yang, C. F., & Zhu, W. Y. (2011). Phylogenetic analysis of methanogens in the pig feces. *Curr Microbiol*, *62*(5), 1386-1389. https://doi.org/10.1007/s00284-011-9873-9

Martínez-Álvaro, M., Auffret, M. D., Stewart, R. D., Dewhurst, R. J., Duthie, C.-A., Rooke, J. A., Wallace, R. J., Shih, B., Freeman, T. C., Watson, M., & Roehe, R. (2020). Identification of Complex Rumen Microbiome Interaction Within Diverse Functional Niches as Mechanisms Affecting the Variation of Methane Emissions in Bovine [Original Research]. *Front Microbiol*, *11*(659). https://doi.org/10.3389/fmicb.2020.00659

Mathur, R., Amichai, M., Chua, K. S., Mirocha, J., Barlow, G. M., & Pimentel, M. (2013). Methane and hydrogen positivity on breath test is associated with greater body mass index and body fat. *J Clin Endocrinol Metab*, *98*(4), E698-702. https://doi.org/10.1210/jc.2012-3144

Maus, I., Wibberg, D., Stantscheff, R., Cibis, K., Eikmeyer, F.-G., König, H., Pühler, A., & Schlüter, A. (2013). Complete genome sequence of the hydrogenotrophic Archaeon Methanobacterium sp. Mb1 isolated from a production-scale biogas plant. *Journal of Biotechnology*, *168*(4), 734-736. https://doi.org/https://doi.org/10.1016/j.jbiotec.2013.10.013

May-Collado, L. J., Kilpatrick, C. W., & Agnarsson, I. (2015). Mammals from 'down under': a multi-gene species-level phylogeny of marsupial mammals (Mammalia, Metatheria). *PeerJ*, *3*, e805. https://doi.org/10.7717/peerj.805

Maya-Lucas, O., Murugesan, S., Nirmalkar, K., Alcaraz, L. D., Hoyo-Vadillo, C., Pizano-Zarate, M. L., & Garcia-Mena, J. (2018). The gut microbiome of Mexican children affected by obesity. *Anaerobe*. https://doi.org/10.1016/j.anaerobe.2018.10.009

McDonald, K. L., & Webb, R. I. (2011). Freeze substitution in 3 hours or less. *Journal of Microscopy*, *243*(3), 227-233. https://doi.org/https://doi.org/10.1111/j.1365-2818.2011.03526.x

McKay, L. F., Eastwood, M. A., & Brydon, W. G. (1985). Methane excretion in man--a study of breath, flatus, and faeces. *Gut*, *26*(1), 69-74. https://gut.bmj.com/content/gutjnl/26/1/69.full.pdf

McKenzie, V. J., Song, S. J., Delsuc, F., Prest, T. L., Oliverio, A. M., Korpita, T. M., Alexiev, A., Amato, K. R., Metcalf, J. L., Kowalewski, M., Avenant, N. L., Link, A., Di Fiore, A., Seguin-Orlando, A., Feh, C., Orlando, L., Mendelson, J. R., Sanders, J., & Knight, R. (2017). The Effects of Captivity on the Mammalian Gut Microbiome. *Integrative and comparative biology*, *57*(4), 690-704. https://doi.org/10.1093/icb/icx090

Meredith, R. W., Westerman, M., & Springer, M. S. (2009). A phylogeny of Diprotodontia (Marsupialia) based on sequences for five nuclear genes. *Molecular Phylogenetics and Evolution*, *51*(3), 554-571. https://doi.org/https://doi.org/10.1016/j.ympev.2009.02.009

Mi, J., Peng, H., Wu, Y., Wang, Y., & Liao, X. (2019). Diversity and community of methanogens in the large intestine of finishing pigs. *BMC Microbiol*, *19*(1), 83. https://doi.org/10.1186/s12866-019-1459-x

Miller, T. L., & Lin, C. (2002). Description of Methanobrevibacter gottschalkii sp. nov., Methanobrevibacter thaueri sp. nov., Methanobrevibacter woesei sp. nov. and Methanobrevibacter wolinii sp. nov. *Int J Syst Evol Microbiol*, *52*(Pt 3), 819-822. https://doi.org/10.1099/00207713-52-3-819

Miller, T. L., & Wolin, M. J. (1985). Methanosphaera stadtmaniae gen. nov., sp. nov.: a species that forms methane by reducing methanol with hydrogen. *Arch Microbiol*, *141*(2), 116-122. https://doi.org/10.1007/bf00423270

Miller, T. L., Wolin, M. J., Conway de Macario, E., & Macario, A. J. (1982). Isolation of Methanobrevibacter smithii from human feces. *Appl Environ Microbiol*, *43*(1), 227-232. https://doi.org/10.1128/aem.43.1.227-232.1982

Miller, T. L., Wolin, M. J., & Kusel, E. A. (1986). Isolation and characterization of methanogens from animal feces. *Systematic and Applied Microbiology*, *8*(3), 234-238. https://doi.org/https://doi.org/10.1016/S0723-2020(86)80084-4

Million, M., Angelakis, E., Maraninchi, M., Henry, M., Giorgi, R., Valero, R., Vialettes, B., & Raoult, D. (2013). Correlation between body mass index and gut concentrations of Lactobacillus reuteri, Bifidobacterium animalis, Methanobrevibacter smithii and Escherichia coli. *International Journal of Obesity*, *37*(11), 1460-1466. https://doi.org/10.1038/ijo.2013.20

Million, M., Maraninchi, M., Henry, M., Armougom, F., Richet, H., Carrieri, P., Valero, R., Raccah, D., Vialettes, B., & Raoult, D. (2012). Obesity-associated gut microbiota is enriched in Lactobacillus reuteri and depleted in Bifidobacterium animalis and Methanobrevibacter smithii. *International Journal of Obesity*, *36*(6), 817-825. https://doi.org/10.1038/ijo.2011.153

Million, M., Tidjani Alou, M., Khelaifia, S., Bachar, D., Lagier, J.-C., Dione, N., Brah, S., Hugon, P., Lombard, V., Armougom, F., Fromonot, J., Robert, C., Michelle, C., Diallo, A., Fabre, A., Guieu, R., Sokhna, C., Henrissat, B., Parola, P., & Raoult, D. (2016). Increased Gut Redox and Depletion of Anaerobic and Methanogenic Prokaryotes in Severe Acute Malnutrition. *Sci Rep*, *6*, 26051-26051. https://doi.org/10.1038/srep26051

Miquel, S., Peyretaillade, E., Claret, L., de Vallée, A., Dossat, C., Vacherie, B., Zineb el, H., Segurens, B., Barbe, V., Sauvanet, P., Neut, C., Colombel, J. F., Medigue, C., Mojica, F. J., Peyret, P., Bonnet, R., & Darfeuille-Michaud, A. (2010). Complete genome sequence of

Crohn's disease-associated adherent-invasive E. coli strain LF82. *Plos One*, *5*(9). https://doi.org/10.1371/journal.pone.0012714

Mira-Pascual, L., Cabrera-Rubio, R., Ocon, S., Costales, P., Parra, A., Suarez, A., Moris, F., Rodrigo, L., Mira, A., & Collado, M. C. (2015). Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers. *J Gastroenterol*, *50*(2), 167-179. https://doi.org/10.1007/s00535-014-0963-x

MM, O. D., Harris, H. M., Jeffery, I. B., Claesson, M. J., Younge, B., PW, O. T., & Ross, R. P. (2013). The core faecal bacterial microbiome of Irish Thoroughbred racehorses. *Lett Appl Microbiol*, *57*(6), 492-501. https://doi.org/10.1111/lam.12137

Moissl-Eichinger, C., Probst, A. J., Birarda, G., Auerbach, A., Koskinen, K., Wolf, P., & Holman, H.-Y. N. (2017). Human age and skin physiology shape diversity and abundance of Archaea on skin. *Sci Rep*, *7*, 4039. https://doi.org/10.1038/s41598-017-04197-4

Motta, J. P., Allain, T., Green-Harrison, L. E., Groves, R. A., Feener, T., Ramay, H., Beck, P. L., Lewis, I. A., Wallace, J. L., & Buret, A. G. (2018). Iron Sequestration in Microbiota Biofilms As A Novel Strategy for Treating Inflammatory Bowel Disease. *Inflamm Bowel Dis*, *24*(7), 1493-1502. https://doi.org/10.1093/ibd/izy116

Naftalin, R., & Pedley, K. (1995). The sodium concentration of lateral intercellular spaces. *The Journal of membrane biology*, *147*(1), 105-107.

Nakamura, N., Lin, H. C., McSweeney, C. S., Mackie, R. I., & Gaskins, H. R. (2010). Mechanisms of microbial hydrogen disposal in the human colon and implications for health and disease. *Annu Rev Food Sci Technol*, *1*, 363-395. https://doi.org/10.1146/annurev.food.102308.124101

Nakano, Y., Suzuki, N., Yoshida, Y., Nezu, T., Yamashita, Y., & Koga, T. (2000). Thymidine diphosphate-6-deoxy-L-lyxo-4-hexulose reductase synthesizing dTDP-6-deoxy-L-talose from Actinobacillus actinomycetemcomitans. *J Biol Chem*, *275*(10), 6806-6812. https://doi.org/10.1074/jbc.275.10.6806

Nakashita, H., Watanabe, K., Hara, O., Hidaka, T., & Seto, H. (1997). Studies on the biosynthesis of bialaphos. Biochemical mechanism of C-P bond formation: discovery of phosphonopyruvate decarboxylase which catalyzes the formation of phosphonoacetaldehyde from phosphonopyruvate. *J Antibiot (Tokyo)*, *50*(3), 212-219. https://www.jstage.jst.go.jp/article/antibiotics1968/50/3/50_3_212/_pdf

Nam, Y. D., Chang, H. W., Kim, K. H., Roh, S. W., Kim, M. S., Jung, M. J., Lee, S. W., Kim, J. Y., Yoon, J. H., & Bae, J. W. (2008). Bacterial, archaeal, and eukaryal diversity in the intestines of Korean people. *J Microbiol*, *46*(5), 491-501. https://doi.org/10.1007/s12275-008-0199-7

Namasivayam, S., Kauffman, K. D., McCulloch, J. A., Yuan, W., Thovarai, V., Mittereder, L. R., Trinchieri, G., Barber, D. L., & Sher, A. (2019). Correlation between Disease Severity and the Intestinal Microbiome in Mycobacterium tuberculosis-Infected Rhesus Macaques. *mBio*, *10*(3). https://doi.org/10.1128/mBio.01018-19

Nava, G. M., Carbonero, F., Croix, J. A., Greenberg, E., & Gaskins, H. R. (2012). Abundance and diversity of mucosa-associated hydrogenotrophic microbes in the healthy human colon. *Isme Journal*, *6*(1), 57-70. https://doi.org/10.1038/ismej.2011.90

Nayfach, S., Shi, Z. J., Seshadri, R., Pollard, K. S., & Kyrpides, N. C. (2019). New insights from uncultivated genomes of the global human gut microbiome. *Nature*, *568*(7753), 505-510. https://doi.org/10.1038/s41586-019-1058-x

NCBI. (2021). *PubChem Annotation Record for METHANE, Source: Hazardous Substances Data Bank (HSDB)*. Retrieved July 16 from

Nehmé, B., Gilbert, Y., Létourneau, V., Forster, R. J., Veillette, M., Villemur, R., & Duchaine, C. (2009). Culture-Independent Characterization of Archaeal Biodiversity in Swine Confinement Building Bioaerosols [10.1128/AEM.00726-09]. *Applied and Environmental Microbiology*, *75*(17), 5445. http://aem.asm.org/content/75/17/5445.abstract

Nguyen-Hieu, T., Khelaifia, S., Aboudharam, G., & Drancourt, M. (2013). Methanogenic archaea in subgingival sites: a review. *Apmis*, *121*(6), 467-477. https://doi.org/10.1111/apm.12015

Nishino, K., Nishida, A., Inoue, R., Kawada, Y., Ohno, M., Sakai, S., Inatomi, O., Bamba, S., Sugimoto, M., Kawahara, M., Naito, Y., & Andoh, A. (2018). Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease. *J Gastroenterol*, *53*(1), 95-106. https://doi.org/10.1007/s00535-017-1384-4

Nkamga, V. D., Henrissat, B., & Drancourt, M. (2017). Archaea: Essential inhabitants of the human digestive microbiota. *Human Microbiome Journal*, *3*, 1-8. https://doi.org/https://doi.org/10.1016/j.humic.2016.11.005

Nkamga, V. D., Lotte, R., Chirio, D., Lonjon, M., Roger, P. M., Drancourt, M., & Ruimy, R. (2018). Methanobrevibacter oralis detected along with Aggregatibacter actinomycetemcomitans in a series of community-acquired brain abscesses. *Clin Microbiol Infect*, *24*(2), 207-208. https://doi.org/10.1016/j.cmi.2017.08.021

Nkamga, V. D., Lotte, R., Roger, P. M., Drancourt, M., & Ruimy, R. (2016). Methanobrevibacter smithii and Bacteroides thetaiotaomicron cultivated from a chronic paravertebral muscle abscess. *Clinical Microbiology and Infection*, *22*(12), 1008-1009. https://doi.org/https://doi.org/10.1016/j.cmi.2016.09.007

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res*, *27*(5), 824-834. https://doi.org/10.1101/gr.213959.116

O'Keefe, S. J., Chung, D., Mahmoud, N., Sepulveda, A. R., Manafe, M., Arch, J., Adada, H., & van der Merwe, T. (2007). Why do African Americans get more colon cancer than Native Africans? *J Nutr*, *137*(1 Suppl), 175s-182s. https://doi.org/10.1093/jn/137.1.175S

Okonechnikov, K., Golosova, O., Fursov, M., & team, t. U. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, *28*(8), 1166-1167. https://doi.org/10.1093/bioinformatics/bts091

Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, *11*(12), 2864-2868. https://doi.org/10.1038/ismej.2017.126

Osawa, R., Blanshard, W., & Ocallaghan, P. (1993). Microbiological Studies of the Intestinal Microflora of the Koala, Phascolarctos-Cinereus .2. Pap, a Special Maternal Feces Consumed by Juvenile Koalas. *Australian Journal of Zoology*, *41*(6), 611-620. https://doi.org/https://doi.org/10.1071/ZO9930611

Osawa, R., Rainey, F., Fujisawa, T., Lang, E., Busse, H. J., Walsh, T. P., & Stackebrandt, E. (1995). Lonepinella koalarum gen. nov., sp. nov., a New Tannin-Protein Complex Degrading Bacterium. *Systematic and Applied Microbiology*, *18*(3), 368-373. https://doi.org/https://doi.org/10.1016/S0723-2020(11)80430-3

Ouwerkerk, D., Maguire, A., McMillen, L., & Klieve, A. (2009). Hydrogen utilising bacteria from the forestomach of eastern grey (Macropus giganteus) and red (Macropus rufus) kangaroos. *Animal Production Science*, *49*(11), 1043-1051.

Oxley, A. P., Lanfranconi, M. P., Wurdemann, D., Ott, S., Schreiber, S., McGenity, T. J., Timmis, K. N., & Nogales, B. (2010). Halophilic archaea in the human intestinal mucosa. *Environ Microbiol*, *12*(9), 2398-2410. https://doi.org/10.1111/j.1462-2920.2010.02212.x

Öztürk, M., & Önal, C. (2019). Asparagine 79 is an important amino acid for catalytic activity and substrate specificity of bile salt hydrolase (BSH). *Mol Biol Rep*, *46*(4), 4361-4368. https://doi.org/10.1007/s11033-019-04889-2

Padmanabha, J., Liu, J., Kurekci, C., Denman, S., & McSweeney, C. (2013). A methylotrophic methanogen isolate from the Thermoplasmatales affiliated RCC clade may provide insight into the role of this group in the rumen. Proceedings of the 5th Greenhouse Gases and Animal Agriculture Conference, Dublin.

Panigrahi, P., Sule, M., Sharma, R., Ramasamy, S., & Suresh, C. G. (2014). An improved method for specificity annotation shows a distinct evolutionary divergence among the microbial enzymes of the cholylglycine hydrolase family. *Microbiology*, *160*(6), 1162-1174. https://doi.org/https://doi.org/10.1099/mic.0.077586-0

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*, *25*(7), 1043-1055. https://doi.org/10.1101/gr.186072.114

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P. A., Woodcroft, B. J., Evans, P. N., Hugenholtz, P., & Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol*, *2*(11), 1533-1542. https://doi.org/10.1038/s41564-017-0012-7

Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., Martinez, X., Varela, E., Sarrabayrouse, G., & Machiels, K. (2017). A microbial signature for Crohn's disease. *Gut*, *66*(5), 813-822. https://gut.bmj.com/content/gutjnl/66/5/813.full.pdf

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M. C., Rice, B. L., DuLong, C., Morgan, X. C., Golden, C. D., Quince, C., Huttenhower, C., & Segata, N. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*, *176*(3), 649-662.e620. https://doi.org/10.1016/j.cell.2019.01.001

Peled, Y., Weinberg, D., Hallak, A., & Gilat, T. (1987). Factors affecting methane production in humans. Gastrointestinal diseases and alterations of colonic flora. *Dig Dis Sci*, *32*(3), 267-271.

Penn, W. D., Harrington, H. R., Schlebach, J. P., & Mukhopadhyay, S. (2020). Regulators of Viral Frameshifting: More Than RNA Influences Translation Events. *Annual Review of Virology*, *7*(1), 219-238. https://doi.org/10.1146/annurev-virology-012120-101548

Peters, B. A., Wilson, M., Moran, U., Pavlick, A., Izsak, A., Wechter, T., Weber, J. S., Osman, I., & Ahn, J. (2019). Relating the gut metagenome and metatranscriptome to immunotherapy responses in melanoma patients. *Genome Med*, *11*(1), 61. https://doi.org/10.1186/s13073-019-0672-4

Petersen, A. M., Nielsen, E. M., Litrup, E., Brynskov, J., Mirsepasi, H., & Krogfelt, K. A. (2009). A phylogenetic group of Escherichia coli associated with active left-sided inflammatory bowel disease. *BMC Microbiol*, *9*, 171. https://doi.org/10.1186/1471-2180-9-171

Peyrin-Biroulet, L., Williet, N., & Cacoub, P. (2015). Guidelines on the diagnosis and treatment of iron deficiency across indications: a systematic review. *Am J Clin Nutr*, *102*(6), 1585-1594. https://doi.org/10.3945/ajcn.114.103366

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*, *29*(9), e45-e45. https://www.ncbi.nlm.nih.gov/pubmed/11328886

Philips, A., Stolarek, I., Kuczkowska, B., Juras, A., Handschuh, L., Piontek, J., Kozlowski, P., & Figlerowicz, M. (2017). Comprehensive analysis of microorganisms accompanying human archaeological remains. *Gigascience*, *6*(7), 1-13. https://doi.org/10.1093/gigascience/gix044

Picchianti-Diamanti, A., Panebianco, C., Salemi, S., Sorgi, M. L., Di Rosa, R., Tropea, A., Sgrulletti, M., Salerno, G., Terracciano, F., D'Amelio, R., Lagana, B., & Pazienza, V. (2018). Analysis of Gut Microbiota in Rheumatoid Arthritis Patients: Disease-Related Dysbiosis and Modifications Induced by Etanercept. *Int J Mol Sci*, *19*(10). https://doi.org/10.3390/ijms19102938

Pieper, H. J., Oplustil, E., & Barth, G. (1980). Reduction of methanol formation during alcoholic fermentation. *Biotechnology Letters*, *2*(9), 391-396. https://doi.org/10.1007/BF00144243

Pietilä, M. K., Laurinmäki, P., Russell, D. A., Ko, C.-C., Jacobs-Sera, D., Hendrix, R. W., Bamford, D. H., & Butcher, S. J. (2013). Structure of the archaeal head-tailed virus HSTV-1 completes the HK97 fold story. *Proceedings of the National Academy of Sciences*, *110*(26), 10604-10609. https://doi.org/10.1073/pnas.1303047110

Pimentel, M., Lin, H. C., Enayati, P., van den Burg, B., Lee, H. R., Chen, J. H., Park, S., Kong, Y., & Conklin, J. (2006). Methane, a gas produced by enteric bacteria, slows intestinal transit and augments small intestinal contractile activity. *Am J Physiol Gastrointest Liver Physiol*, *290*(6), G1089-1095. https://doi.org/10.1152/ajpgi.00574.2004

Pimentel, M., Mayer, A. G., Park, S., Chow, E. J., Hasan, A., & Kong, Y. (2003). Methane production during lactulose breath test is associated with gastrointestinal disease presentation. *Dig Dis Sci*, *48*(1), 86-92.

Pimentel, M., Saad, R. J., Long, M. D., & Rao, S. S. C. (2020). ACG Clinical Guideline: Small Intestinal Bacterial Overgrowth. *Official journal of the American College of Gastroenterology | ACG*, *115*(2), 165-178. https://doi.org/10.14309/ajg.0000000000000501

Pique, J. M., Pallares, M., Cuso, E., Vilar-Bonet, J., & Gassull, M. A. (1984). Methane production and colon cancer. *Gastroenterology*, *87*(3), 601-605.

Pisa, K. Y., Weidner, C., Maischak, H., Kavermann, H., & Muller, V. (2007). The coupling ion in the methanoarchaeal ATP synthases: H(+) vs. Na(+) in the A(1)A(o) ATP synthase from the archaeon Methanosarcina mazei Go1. *FEMS Microbiol Lett*, *277*(1), 56-63. https://doi.org/10.1111/j.1574-6968.2007.00939.x

Pollock, J., Muwonge, A., Hutchings, M. R., Mainda, G., Bronsvoort, B. M., Gally, D. L., & Corbishley, A. (2020). Resistance to change: AMR gene dynamics on a commercial pig farm with high antimicrobial usage. *Sci Rep*, *10*(1), 1708. https://doi.org/10.1038/s41598-020-58659-3

Ponziani, F. R., Picca, A., Marzetti, E., Calvani, R., Conta, G., Del Chierico, F., Capuani, G., Faccia, M., Fianchi, F., Funaro, B., Josè Coelho-Junior, H., Petito, V., Rinninella, E., Paroni Sterbini, F., Reddel, S., Vernocchi, P., Cristina Mele, M., Miccheli, A., Putignani, L., Sanguinetti, M., Pompili, M., Gasbarrini, A., & group, t. G. s. (2021). Characterization of the gut-liver-muscle axis in cirrhotic patients with sarcopenia. *Liver International*, *41*(6), 1320-1334. https://doi.org/https://doi.org/10.1111/liv.14876

Porter, A. W., & Young, L. Y. (2014). Chapter Five - Benzoyl-CoA, a Universal Biomarker for Anaerobic Degradation of Aromatic Compounds. In S. Sariaslani & G. M. Gadd (Eds.), *Advances in Applied Microbiology* (Vol. 88, pp. 167-203). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-800260-5.00005-X

Poulsen, C. S. (2019). *The effect of sample processing on microbiomes using metagenomics characterization* Technical University of Denmark].

Pozuelo, M., Panda, S., Santiago, A., Mendez, S., Accarino, A., Santos, J., Guarner, F., Azpiroz, F., & Manichanh, C. (2015). Reduction of butyrate- and methane-producing microorganisms in patients with Irritable Bowel Syndrome [Article]. *Sci Rep*, *5*, 12693. https://doi.org/10.1038/srep12693

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *Plos One*, *5*(3), e9490. https://doi.org/10.1371/journal.pone.0009490

Prochazkova, P., Roubalova, R., Dvorak, J., Kreisinger, J., Hill, M., Tlaskalova-Hogenova, H., Tomasova, P., Pelantova, H., Cermakova, M., Kuzma, M., Bulant, J., Bilej, M., Smitka, K., Lambertova, A., Holanova, P., & Papezova, H. (2021). The intestinal microbiota and metabolites in patients with anorexia nervosa. *Gut Microbes*, *13*(1), 1-25. https://doi.org/10.1080/19490976.2021.1902771

Qian, Y., Yang, X., Xu, S., Wu, C., Song, Y., Qin, N., Chen, S. D., & Xiao, Q. (2018). Alteration of the fecal microbiota in Chinese patients with Parkinson's disease. *Brain Behav Immun*, *70*, 194-202. https://doi.org/10.1016/j.bbi.2018.02.016

Qin, W., Song, P., Lin, G., Huang, Y., Wang, L., Zhou, X., Li, S., & Zhang, T. (2020). Gut Microbiota Plasticity Influences the Adaptability of Wild and Domestic Animals in Co-inhabited Areas. *Front Microbiol*, *11*, 125. https://doi.org/10.3389/fmicb.2020.00125

Ramezani, A., Nolin, T. D., Barrows, I. R., Serrano, M. G., Buck, G. A., Regunathan-Shenk, R., West, R. E., Latham, P. S., Amdur, R., & Raj, D. S. (2018). Gut Colonization with Methanogenic Archaea Lowers Plasma Trimethylamine N-oxide Concentrations in Apolipoprotein e−/− Mice. *Sci Rep*, *8*(1), 14752. https://doi.org/10.1038/s41598-018-33018-5

Rasmussen, R. (2001). Quantification on the LightCycler. In: Meuer S, Wittwer C, Nakagawara K-I (eds). Rapid Cycle Real-Time PCR. *Springer: Berlin, Heidelberg*, pp 21–34.

Rebouche, C. J., & Chenard, C. A. (1991). Metabolic fate of dietary carnitine in human adults: identification and quantification of urinary and fecal metabolites. *J Nutr*, *121*(4), 539-546. https://doi.org/10.1093/jn/121.4.539

Revilla, I., & González-SanJosé, M. L. (1998). Methanol release during fermentation of red grapes treated with pectolytic enzymes. *Food Chemistry*, *63*(3), 307-312. https://doi.org/https://doi.org/10.1016/S0308-8146(98)00049-1

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, *16*(6), 276-277. https://doi.org/https://doi.org/10.1016/S0168-9525(00)02024-2

Richardson, A. R., & Stojiljkovic, I. (1999). HmbR, a hemoglobin-binding outer membrane protein of Neisseria meningitidis, undergoes phase variation. *J Bacteriol*, *181*(7), 2067-2074. https://doi.org/10.1128/jb.181.7.2067-2074.1999

Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A*, *106*(45), 19126-19131. https://doi.org/10.1073/pnas.0906412106

Rieu-Lesme, F., Delbes, C., & Sollelis, L. (2005). Recovery of partial 16S rDNA sequences suggests the presence of Crenarchaeota in the human digestive ecosystem. *Curr Microbiol*, *51*(5), 317-321. https://doi.org/10.1007/s00284-005-0036-8

Robichaux, M., Howell, M., & Boopathy, R. (2003). Methanogenic activity in human periodontal pocket. *Curr Microbiol*, *46*(1), 53-58. https://doi.org/10.1007/s00284-002-3807-5

Roncal, C., Martínez-Aguilar, E., Orbe, J., Ravassa, S., Fernandez-Montero, A., Saenz-Pipaon, G., Ugarte, A., Estella-Hermoso de Mendoza, A., Rodriguez, J. A., Fernández-Alonso, S., Fernández-Alonso, L., Oyarzabal, J., & Paramo, J. A. (2019). Trimethylamine-N-Oxide (TMAO) Predicts Cardiovascular Mortality in Peripheral Artery Disease. *Sci Rep*, *9*(1), 15580. https://doi.org/10.1038/s41598-019-52082-z

Rovira Sanz, P. (2017). *Impact of antibiotic use on resistance in beef feedlot and dairy cattle* Colorado State University].

Salgado-Flores, A., Tveit, A. T., Wright, A.-D., Pope, P. B., & Sundset, M. A. (2019). Characterization of the cecum microbiome from wild and captive rock ptarmigans indigenous to Arctic Norway. *Plos One*, *14*(3), e0213503. https://doi.org/10.1371/journal.pone.0213503

Samuel, B. S., Hansen, E. E., Manchester, J. K., Coutinho, P. M., Henrissat, B., Fulton, R., Latreille, P., Kim, K., Wilson, R. K., & Gordon, J. I. (2007). Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(25), 10643-10648. https://doi.org/10.1073/pnas.0704189104

Sandborn, W. J. (2009). How to avoid treating irritable bowel syndrome with biologic therapy for inflammatory bowel disease. *Dig Dis*, *27 Suppl 1*, 80-84. https://doi.org/10.1159/000268125

Sansonetti, P. J. (2004). War and peace at mucosal surfaces. *Nat Rev Immunol*, *4*(12), 953-964. https://doi.org/10.1038/nri1499

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., Dlugokencky, E. J., Etiope, G., Bastviken, D., Houweling, S., Janssens-Maenhout, G., Tubiello, F. N., Castaldi, S., Jackson, R. B., Alexe, M., Arora, V. K., Beerling, D. J., Bergamaschi, P., Blake, D. R., Brailsford, G., Brovkin, V., Bruhwiler, L., Crevoisier, C., Crill, P., Covey, K., Curry, C., Frankenberg, C., Gedney, N., Höglund-Isaksson, L., Ishizawa, M., Ito, A., Joos, F., Kim, H. S., Kleinen, T., Krummel, P., Lamarque, J. F., Langenfelds, R., Locatelli, R., Machida, T., Maksyutov, S., McDonald, K. C., Marshall, J., Melton, J. R., Morino, I., Naik, V., O'Doherty, S., Parmentier, F. J. W., Patra, P. K., Peng, C., Peng, S., Peters, G. P., Pison, I., Prigent, C., Prinn, R., Ramonet, M., Riley, W. J., Saito, M., Santini, M., Schroeder, R., Simpson, I. J., Spahni, R., Steele, P., Takizawa, A., Thornton, B. F., Tian, H., Tohjima, Y., Viovy, N., Voulgarakis, A., van Weele, M., van der Werf, G. R., Weiss, R., Wiedinmyer, C., Wilton, D. J., Wiltshire, A., Worthy, D., Wunch, D., Xu, X., Yoshida, Y., Zhang, B., Zhang, Z., & Zhu, Q. (2016). The global methane budget 2000–2012. *Earth Syst. Sci. Data*, *8*(2), 697-751. https://doi.org/10.5194/essd-8-697-2016

Scanlan, P. D., Shanahan, F., & Marchesi, J. R. (2008). Human methanogen diversity and incidence in healthy and diseased colonic groups using mcrA gene analysis. *Bmc Microbiology*, *8*, Article 79. https://doi.org/10.1186/1471-2180-8-79

Schweiger, G., Dutscho, R., & Buckel, W. (1987). Purification of 2-hydroxyglutaryl-CoA dehydratase from Acidaminococcus fermentans. An iron-sulfur protein. *Eur J Biochem*, *169*(2), 441-448. https://doi.org/10.1111/j.1432-1033.1987.tb13631.x

Schwiertz, A., Taras, D., Schafer, K., Beijer, S., Bos, N. A., Donus, C., & Hardt, P. D. (2010). Microbiota and SCFA in lean and overweight healthy subjects. *Obesity (Silver Spring)*, *18*(1), 190-195. https://doi.org/10.1038/oby.2009.167

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068-2069. https://doi.org/10.1093/bioinformatics/btu153

Segal, I., Walker, A. R., Lord, S., & Cummings, J. H. (1988). Breath methane and large bowel cancer risk in contrasting African populations. *Gut*, *29*(5), 608-613. https://www.ncbi.nlm.nih.gov/pubmed/3396948

Self, W. T., Grunden, A. M., Hasona, A., & Shanmugam, K. T. (2001). Molybdate transport☆☆Florida Agricultural Experiment Station Journal Series No. R-08044. *Research in Microbiology*, *152*(3), 311-321. https://doi.org/https://doi.org/10.1016/S0923-2508(01)01202-5

Senčilo, A., Jacobs-Sera, D., Russell, D. A., Ko, C. C., Bowman, C. A., Atanasova, N. S., Österlund, E., Oksanen, H. M., Bamford, D. H., Hatfull, G. F., Roine, E., & Hendrix, R. W. (2013). Snapshot of haloarchaeal tailed virus genomes. *RNA Biol*, *10*(5), 803-816. https://doi.org/10.4161/rna.24045

Shah, A., Morrison, M., & Holtmann, G. (2018). A novel treatment for patients with constipation: Dawn of a new age for translational microbiome research? *Indian Journal of Gastroenterology*, *37*(5), 388-391. https://doi.org/10.1007/s12664-018-0912-3

Shao, Z., Blodgett, J. A., Circello, B. T., Eliot, A. C., Woodyer, R., Li, G., van der Donk, W. A., Metcalf, W. W., & Zhao, H. (2008). Biosynthesis of 2-hydroxyethylphosphonate, an unexpected intermediate common to multiple phosphonate biosynthetic pathways. *J Biol Chem*, *283*(34), 23161-23168. https://doi.org/10.1074/jbc.M801788200

Shi, W., Moon, C. D., Leahy, S. C., Kang, D., Froula, J., Kittelmann, S., Fan, C., Deutsch, S., Gagic, D., Seedorf, H., Kelly, W. J., Atua, R., Sang, C., Soni, P., Li, D., Pinares-Patiño, C. S., McEwan, J. C., Janssen, P. H., Chen, F., Visel, A., Wang, Z., Attwood, G. T., & Rubin, E. M. (2014). Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res*, *24*(9), 1517-1525. https://doi.org/10.1101/gr.168245.113

Shiffman, M., Soo, R., Dennis, P., Morrison, M., Tyson, G., & Hugenholtz, P. (2017). Gene and genome-centric analyses of koala and wombat fecal microbiomes point to metabolic specialization for Eucalyptus digestion. *PeerJ*, *5*, e4075-e4075. https://doi.org/10.7717/peerj.4075

Shimshoni, E., Ghini, V., Solomonov, I., Luchinat, C., Sagi, I., & Turano, P. (2020). Integrated metabolomics and proteomics of symptomatic and early pre-symptomatic states of colitis. *bioRxiv*, 2020.2003.2022.002196. https://doi.org/10.1101/2020.03.22.002196

Sieber, J. R., McInerney, M. J., & Gunsalus, R. P. (2012). Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Annu Rev Microbiol*, *66*, 429-452. https://doi.org/10.1146/annurev-micro-090110-102844

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, *7*, 539. https://doi.org/10.1038/msb.2011.75

Song, M., Yang, Q., Zhang, F., Chen, L., Su, H., Yang, X., He, H., Liu, F., Zheng, J., Ling, M., Lai, X., Zhu, X., Wang, L., Gao, P., Shu, G., Jiang, Q., & Wang, S. (2020). Hyodeoxycholic acid (HDCA) suppresses intestinal epithelial cell proliferation through FXR-PI3K/AKT pathway, accompanied by alteration of bile acids metabolism profiles induced by gut bacteria. *The FASEB Journal*, *34*(5), 7103-7117. https://doi.org/https://doi.org/10.1096/fj.201903244R

Sorokin, D. Y., Makarova, K. S., Abbas, B., Ferrer, M., Golyshin, P. N., Galinski, E. A., Ciordia, S., Mena, M. C., Merkel, A. Y., Wolf, Y. I., van Loosdrecht, M. C. M., & Koonin, E. V. (2017). Discovery of extremely halophilic, methyl-reducing euryarchaea provides insights into the evolutionary origin of methanogenesis. *Nat Microbiol*, *2*, 17081. https://doi.org/10.1038/nmicrobiol.2017.81

Spain, J. C., & Gibson, D. T. (1991). Pathway for Biodegradation of p-Nitrophenol in a Moraxella sp. *Appl Environ Microbiol*, *57*(3), 812-819. https://doi.org/10.1128/aem.57.3.812-819.1991

Spencer, M. D., Hamp, T. J., Reid, R. W., Fischer, L. M., Zeisel, S. H., & Fodor, A. A. (2011). Association between composition of the human gastrointestinal microbiome and development of fatty liver with choline deficiency. *Gastroenterology*, *140*(3), 976-986. https://doi.org/10.1053/j.gastro.2010.11.049

Spring, K. R. (1998). Routes and mechanism of fluid transport by epithelia. *Annu Rev Physiol*, *60*, 105-119. https://doi.org/10.1146/annurev.physiol.60.1.105

St-Pierre, B., & Wright, A. D. (2013). Diversity of gut methanogens in herbivorous animals. *Animal*, *7 Suppl 1*, 49-56. https://doi.org/10.1017/s1751731112000912

Staley, C., Weingarden, A. R., Khoruts, A., & Sadowsky, M. J. (2017). Interaction of gut microbiota with bile acid metabolism and its influence on disease states. *Appl Microbiol Biotechnol*, *101*(1), 47-64. https://doi.org/10.1007/s00253-016-8006-6

Stams, A. J., & Plugge, C. M. (2009). Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat Rev Microbiol*, *7*(8), 568-577. https://doi.org/10.1038/nrmicro2166

Stein, J., Hartmann, F., & Dignass, A. U. (2010). Diagnosis and management of iron deficiency anemia in patients with IBD. *Nat Rev Gastroenterol Hepatol*, *7*(11), 599-610. https://doi.org/10.1038/nrgastro.2010.151

Stewart, R. D., Auffret, M. D., Warr, A., Walker, A. W., Roehe, R., & Watson, M. (2019). Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nature Biotechnology*, *37*(8), 953-961. https://doi.org/10.1038/s41587-019-0202-3

Sturm, M., Schroeder, C., & Bauer, P. (2016). SeqPurge: highly-sensitive adapter trimming for paired-end NGS data. *BMC Bioinformatics*, *17*(1), 208. https://doi.org/10.1186/s12859-016-1069-7

Suri, J., Kataria, R., Malik, Z., Parkman, H. P., & Schey, R. (2018). Elevated methane levels in small intestinal bacterial overgrowth suggests delayed small bowel and colonic transit. *Medicine (Baltimore)*, *97*(21), e10554. https://doi.org/10.1097/md.0000000000010554

Svartström, O., Alneberg, J., Terrapon, N., Lombard, V., de Bruijn, I., Malmsten, J., Dalin, A. M., El Muller, E., Shah, P., Wilmes, P., Henrissat, B., Aspeborg, H., & Andersson, A. F. (2017). Ninety-nine de novo assembled genomes from the moose (Alces alces) rumen microbiome provide new insights into microbial plant biomass degradation. *Isme j*, *11*(11), 2538-2551. https://doi.org/10.1038/ismej.2017.108

Takenaka, S., Koshiya, J., Okugawa, S., Takata, A., Murakami, S., & Aoki, K. (2011). Fe-superoxide dismutase and 2-hydroxy-1,4-benzoquinone reductase preclude the auto-oxidation step in 4-aminophenol metabolism by Burkholderia sp. strain AK-5. *Biodegradation*, *22*(1), 1-11. https://doi.org/10.1007/s10532-010-9369-5

Tan, H. Y., Sieo, C. C., Abdullah, N., Liang, J. B., Huang, X. D., & Ho, Y. W. (2011). Effects of condensed tannins from Leucaena on methane production, rumen fermentation and populations of methanogens and protozoa in vitro. *Animal Feed Science and Technology*, *169*(3), 185-193. https://doi.org/https://doi.org/10.1016/j.anifeedsci.2011.07.004

Tan, Z., Yang, T., Wang, Y., Xing, K., Zhang, F., Zhao, X., Ao, H., Chen, S., Liu, J., & Wang, C. (2017). Metagenomic Analysis of Cecal Microbiome Identified Microbiota and Functional Capacities Associated with Feed Efficiency in Landrace Finishing Pigs. *Front Microbiol*, *8*, 1546. https://doi.org/10.3389/fmicb.2017.01546

Tang, W. H., Wang, Z., Levison, B. S., Koeth, R. A., Britt, E. B., Fu, X., Wu, Y., & Hazen, S. L. (2013). Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med*, *368*(17), 1575-1584. https://doi.org/10.1056/NEJMoa1109400

Tang, Y., Zhang, J., Li, J., Lei, X., Xu, D., Wang, Y., Li, C., Li, X., & Mao, Y. (2019). Turnover of bile acids in liver, serum and caecal content by high-fat diet feeding affects hepatic steatosis

in rats. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, *1864*(10), 1293-1304. https://doi.org/https://doi.org/10.1016/j.bbalip.2019.05.016

Tap, J., Derrien, M., Tornblom, H., Brazeilles, R., Cools-Portier, S., Dore, J., Storsrud, S., Le Neve, B., Ohman, L., & Simren, M. (2017). Identification of an Intestinal Microbiota Signature Associated With Severity of Irritable Bowel Syndrome. *Gastroenterology*, *152*(1), 111-123.e118. https://doi.org/10.1053/j.gastro.2016.09.049

Teh, J. J., Berendsen, E. M., Hoedt, E. C., Kang, S., Zhang, J., Zhang, F., Liu, Q., Hamilton, A. L., Wilson-O'Brien, A., Ching, J., Sung, J. J. Y., Yu, J., Ng, S. C., Kamm, M. A., & Morrison, M. (2021). Novel strain-level resolution of Crohn's disease mucosa-associated microbiota via an ex vivo combination of microbe culture and metagenomic sequencing. *The ISME Journal*. https://doi.org/10.1038/s41396-021-00991-1

Tett, A., Huang, K. D., Asnicar, F., Fehlner-Peach, H., Pasolli, E., Karcher, N., Armanini, F., Manghi, P., Bonham, K., Zolfo, M., De Filippis, F., Magnabosco, C., Bonneau, R., Lusingu, J., Amuasi, J., Reinhard, K., Rattei, T., Boulund, F., Engstrand, L., Zink, A., Collado, M. C., Littman, D. R., Eibach, D., Ercolini, D., Rota-Stabelli, O., Huttenhower, C., Maixner, F., & Segata, N. (2019). The Prevotella copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations. *Cell Host Microbe*, *26*(5), 666-679.e667. https://doi.org/10.1016/j.chom.2019.08.018

Thauer, R. K. (1998). Biochemistry of methanogenesis: a tribute to Marjory Stephenson:1998 Marjory Stephenson Prize Lecture. *Microbiology*, *144*(9), 2377-2406. https://doi.org/https://doi.org/10.1099/00221287-144-9-2377

Ticak, T., Hariraju, D., Arcelay, M. B., Arivett, B. A., Fiester, S. E., & Ferguson, D. J., Jr. (2015). Isolation and characterization of a tetramethylammonium-degrading Methanococcoides strain and a novel glycine betaine-utilizing Methanolobus strain. *Arch Microbiol*, *197*(2), 197-209. https://doi.org/10.1007/s00203-014-1043-6

Toxicity, C. o. (2011). CoT Statement on the Effects of Chronic Dietary Exposure to Methanol. *Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment* https://cot.food.gov.uk/sites/default/files/cot/cotstatementmethanol201102revjuly.pdf

Tremlett, H., Fadrosh, D. W., Faruqi, A. A., Hart, J., Roalstad, S., Graves, J., Lynch, S., & Waubant, E. (2016). Gut microbiota composition and relapse risk in pediatric MS: A pilot study. *J Neurol Sci*, *363*, 153-157. https://doi.org/10.1016/j.jns.2016.02.042

Tremlett, H., Fadrosh, D. W., Faruqi, A. A., Zhu, F., Hart, J., Roalstad, S., Graves, J., Lynch, S., Waubant, E., & Centers, t. U. N. o. P. M. (2016). Gut microbiota in early pediatric multiple sclerosis: a case−control study. *European Journal of Neurology*, *23*(8), 1308-1321. https://doi.org/https://doi.org/10.1111/ene.13026

Trent, M. S., Ribeiro, A. A., Lin, S., Cotter, R. J., & Raetz, C. R. (2001). An inner membrane enzyme in Salmonella and Escherichia coli that transfers 4-amino-4-deoxy-L-arabinose to lipid A: induction on polymyxin-resistant mutants and role of a novel lipid-linked donor. *J Biol Chem*, *276*(46), 43122-43131. https://doi.org/10.1074/jbc.M106961200

Triantafyllou, K., Chang, C., & Pimentel, M. (2014). Methanogens, methane and gastrointestinal motility. *J Neurogastroenterol Motil*, *20*(1), 31-40. https://doi.org/10.5056/jnm.2014.20.1.31

USEPA. (2012). US EPA; Estimation Program Interface (EPI) Suite. Ver. 4.1.

Vanderhaeghen, S., Lacroix, C., & Schwab, C. (2015). Methanogen communities in stools of humans of different age and health status and co-occurrence with bacteria. *Fems Microbiology Letters*, *362*(13), fnv092-fnv092. https://doi.org/10.1093/femsle/fnv092

Vendl, C., Clauss, M., Stewart, M., Leggett, K., Hummel, J., Kreuzer, M., & Munn, A. (2015). Decreasing methane yield with increasing food intake keeps daily methane emissions constant in two foregut fermenting marsupials, the western grey kangaroo and red kangaroo. *Journal of Experimental Biology*, *218*(21), 3425-3434. https://doi.org/10.1242/jeb.128165

Verma, R., Verma, A. K., Ahuja, V., & Paul, J. (2010). Real-time analysis of mucosal flora in patients with inflammatory bowel disease in India. *J Clin Microbiol*, *48*(11), 4279-4282. https://doi.org/10.1128/jcm.01360-10

Vierbuchen, T., Bang, C., Rosigkeit, H., Schmitz, R. A., & Heine, H. (2017). The Human-Associated Archaeon Methanosphaera stadtmanae Is Recognized through Its RNA and Induces TLR8-Dependent NLRP3 Inflammasome Activation. *Front Immunol*, *8*, 1535. https://doi.org/10.3389/fimmu.2017.01535

Von Engelhardt, W., Wolter, S., Lawrenz, H., & Hemsley, J. (1978). Production of methane in two non-ruminant herbivores. *Comparative Biochemistry and Physiology Part A: Physiology*, *60*(3), 309-311.

Wallace, R. J., Rooke, J. A., McKain, N., Duthie, C. A., Hyslop, J. J., Ross, D. W., Waterhouse, A., Watson, M., & Roehe, R. (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics*, *16*, 839. https://doi.org/10.1186/s12864-015-2032-0

Wang, C., Li, P., Yan, Q., Chen, L., Li, T., Zhang, W., Li, H., Chen, C., Han, X., Zhang, S., Xu, M., Li, B., Zhang, X., Ni, H., Ma, Y., Dong, B., Li, S., & Liu, S. (2019). Characterization of the Pig Gut Microbiome and Antibiotic Resistome in Industrialized Feedlots in China. *mSystems*, *4*(6). https://doi.org/10.1128/mSystems.00206-19

Wang, H., Yan, Y., Yi, X., Duan, Y., Wang, J., Li, S., Luo, L., Huang, T., Inglis, B., Li, X., Ji, W., Tan, T., & Si, W. (2019). Histopathological Features and Composition of Gut Microbiota in

Rhesus Monkey of Alcoholic Liver Disease. *Front Microbiol*, *10*, 165. https://doi.org/10.3389/fmicb.2019.00165

Wang, Q., Li, F., Liang, B., Liang, Y., Chen, S., Mo, X., Ju, Y., Zhao, H., Jia, H., Spector, T. D., Xie, H., & Guo, R. (2018). A metagenome-wide association study of gut microbiota in asthma in UK adults. *Bmc Microbiology*, *18*(1), 114. https://doi.org/10.1186/s12866-018-1257-x

Wang, X., Xuan, Y., Han, Y., Ding, X., Ye, K., Yang, F., Gao, P., Goff, S. P., & Gao, G. (2019). Regulation of HIV-1 Gag-Pol Expression by Shiftless, an Inhibitor of Programmed -1 Ribosomal Frameshifting. *Cell*, *176*(3), 625-635.e614. https://doi.org/https://doi.org/10.1016/j.cell.2018.12.030

Watkins, A. J., Roussel, E. G., Webster, G., Parkes, R. J., & Sass, H. (2012). Choline and N,N-Dimethylethanolamine as Direct Substrates for Methanogens [10.1128/AEM.01941-12]. *Applied and Environmental Microbiology*, *78*(23), 8298. http://aem.asm.org/content/78/23/8298.abstract

Weaver, G. A., Krause, J. A., Miller, T. L., & Wolin, M. J. (1986). Incidence of methanogenic bacteria in a sigmoidoscopy population: an association of methanogenic bacteria and diverticulosis. *Gut*, *27*(6), 698-704. https://www.ncbi.nlm.nih.gov/pubmed/3721294

Welte, C., & Deppenmeier, U. (2011). Chapter thirteen - Proton Translocation in Methanogens. In A. C. Rosenzweig & S. W. Ragsdale (Eds.), *Methods in Enzymology* (Vol. 494, pp. 257-280). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-385112-3.00013-5

Weng, Y. J., Gan, H. Y., Li, X., Huang, Y., Li, Z. C., Deng, H. M., Chen, S. Z., Zhou, Y., Wang, L. S., Han, Y. P., Tan, Y. F., Song, Y. J., Du, Z. M., Liu, Y. Y., Wang, Y., Qin, N., Bai, Y., Yang, R. F., Bi, Y. J., & Zhi, F. C. (2019). Correlation of diet, microbiota and metabolite networks in inflammatory bowel disease. *J Dig Dis*, *20*(9), 447-459. https://doi.org/10.1111/1751-2980.12795

Werner, T., Wagner, S. J., Martínez, I., Walter, J., Chang, J.-S., Clavel, T., Kisling, S., Schuemann, K., & Haller, D. (2011). Depletion of luminal iron alters the gut microbiota and prevents Crohn's disease-like ileitis. *Gut*, *60*(3), 325-333. https://doi.org/10.1136/gut.2010.216929

White, B. A., & Mackie, R. I. (1997). Gastrointestinal microbiology. Vol. 1, Gastrointestinal ecosystems and fermentations.

Whitford, M. F., Teather, R. M., & Forster, R. J. (2001). Phylogenetic analysis of methanogens from the bovine rumen. *Bmc Microbiology*, *1*(1), 5. https://doi.org/10.1186/1471-2180-1-5

Wolf, J., Asrar, G. R., & West, T. O. (2017). Revised methane emissions factors and spatially distributed annual carbon fluxes for global livestock. *Carbon Balance and Management*, *12*(1), 16. https://doi.org/10.1186/s13021-017-0084-y

Wolff, S. M., Ellison, M. J., Hao, Y., Cockrum, R. R., Austin, K. J., Baraboo, M., Burch, K., Lee, H. J., Maurer, T., Patil, R., Ravelo, A., Taxis, T. M., Truong, H., Lamberson, W. R., Cammack, K. M., & Conant, G. C. (2017). Diet shifts provoke complex and variable changes in the metabolic networks of the ruminal microbiome. *Microbiome*, *5*(1), 60. https://doi.org/10.1186/s40168-017-0274-6

Wood, G. E., Haydock, A. K., & Leigh, J. A. (2003). Function and regulation of the formate dehydrogenase genes of the methanogenic archaeon Methanococcus maripaludis. *J Bacteriol*, *185*(8), 2548-2554. https://doi.org/10.1128/jb.185.8.2548-2554.2003

Woodyer, R. D., Shao, Z., Thomas, P. M., Kelleher, N. L., Blodgett, J. A., Metcalf, W. W., van der Donk, W. A., & Zhao, H. (2006). Heterologous production of fosfomycin and identification of the minimal biosynthetic gene cluster. *Chem Biol*, *13*(11), 1171-1182. https://doi.org/10.1016/j.chembiol.2006.09.007

Worm, P., Müller, N., Plugge, C. M., Stams, A. J., & Schink, B. (2010). Syntrophy in methanogenic degradation. In *(Endo) symbiotic methanogenic archaea* (pp. 143-173). Springer.

Wright, A. D., & Pimm, C. (2003). Improved strategy for presumptive identification of methanogens using 16S riboprinting. *J Microbiol Methods*, *55*(2), 337-349. http://ac.els-cdn.com/S0167701203001696/1-s2.0-S0167701203001696-main.pdf?_tid=0b7f6bf6-a639-11e6-b5d5-00000aacb360&acdnat=1478667641_426d500bec0538251a303d1943483f77

Wyres, K. L., Wick, R. R., Gorrie, C., Jenney, A., Follador, R., Thomson, N. R., & Holt, K. E. (2016). Identification of Klebsiella capsule synthesis loci from whole genome data. *Microb Genom*, *2*(12). https://doi.org/https://doi.org/10.1099/mgen.0.000102

Xu, H., Aurora, R., Rose, G. D., & White, R. H. (1999). Identifying two ancient enzymes in Archaea using predicted secondary structure alignment. *Nat Struct Biol*, *6*(8), 750-754. https://doi.org/10.1038/11525

Xue, Y., Lin, L., Hu, F., Zhu, W., & Mao, S. (2020). Disruption of ruminal homeostasis by malnutrition involved in systemic ruminal microbiota-host interactions in a pregnant sheep model. *Microbiome*, *8*(1), 138. https://doi.org/10.1186/s40168-020-00916-8

Yang, X., Noyes, N. R., Doster, E., Martin, J. N., Linke, L. M., Magnuson, R. J., Yang, H., Geornaras, I., Woerner, D. R., Jones, K. L., Ruiz, J., Boucher, C., Morley, P. S., & Belk, K. E. (2016). Use of Metagenomic Shotgun Sequencing Technology To Detect Foodborne Pathogens within the Microbiome of the Beef Production Chain. *Appl Environ Microbiol*, *82*(8), 2433-2443. https://doi.org/10.1128/aem.00078-16

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*, 134. https://doi.org/10.1186/1471-2105-13-134

Ye, Z., Zhang, N., Wu, C., Zhang, X., Wang, Q., Huang, X., Du, L., Cao, Q., Tang, J., Zhou, C., Hou, S., He, Y., Xu, Q., Xiong, X., Kijlstra, A., Qin, N., & Yang, P. (2018). A metagenomic study of the gut microbiome in Behcet's disease. *Microbiome*, *6*(1), 135. https://doi.org/10.1186/s40168-018-0520-6

Yu, J., Feng, Q., Wong, S. H., Zhang, D., Liang, Q. Y., Qin, Y., Tang, L., Zhao, H., Stenvang, J., Li, Y., Wang, X., Xu, X., Chen, N., Wu, W. K., Al-Aama, J., Nielsen, H. J., Kiilerich, P., Jensen, B. A., Yau, T. O., Lan, Z., Jia, H., Li, J., Xiao, L., Lam, T. Y., Ng, S. C., Cheng, A. S., Wong, V. W., Chan, F. K., Xu, X., Yang, H., Madsen, L., Datz, C., Tilg, H., Wang, J., Brünner, N., Kristiansen, K., Arumugam, M., Sung, J. J., & Wang, J. (2017). Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut*, *66*(1), 70-78. https://doi.org/10.1136/gutjnl-2015-309800

Yu, X., Doroghazi, J. R., Janga, S. C., Zhang, J. K., Circello, B., Griffin, B. M., Labeda, D. P., & Metcalf, W. W. (2013). Diversity and abundance of phosphonate biosynthetic genes in nature. *Proceedings of the National Academy of Sciences*, *110*(51), 20759. https://doi.org/10.1073/pnas.1315107110

Yu, Z., & Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques*, *36*(5), 808-812.

Zaheer, R., Lakin, S. M., Polo, R. O., Cook, S. R., Larney, F. J., Morley, P. S., Booker, C. W., Hannon, S. J., Van Domselaar, G., Read, R. R., & McAllister, T. A. (2019). Comparative diversity of microbiomes and Resistomes in beef feedlots, downstream environments and urban sewage influent. *Bmc Microbiology*, *19*(1), 197. https://doi.org/10.1186/s12866-019-1548-x

Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D., Weinberger, A., Ben-Yacov, O., Lador, D., Avnit-Sagi, T., Lotan-Pompan, M., Suez, J., Mahdi, J. A., Matot, E., Malka, G., Kosower, N., Rein, M., Zilberman-Schapira, G., Dohnalová, L., Pevsner-Fischer, M., Bikovsky, R., Halpern, Z., Elinav, E., & Segal, E. (2015). Personalized Nutrition by Prediction of Glycemic Responses. *Cell*, *163*(5), 1079-1094. https://doi.org/10.1016/j.cell.2015.11.001

Zeisel, S. H., Wishnok, J. S., & Blusztajn, J. K. (1983). Formation of methylamines from ingested choline and lecithin. *J Pharmacol Exp Ther*, *225*(2), 320-324. https://jpet.aspetjournals.org/content/jpet/225/2/320.full.pdf

Zeller, G., Tap, J., Voigt, A. Y., Sunagawa, S., Kultima, J. R., Costea, P. I., Amiot, A., Böhm, J., Brunetti, F., Habermann, N., Hercog, R., Koch, M., Luciani, A., Mende, D. R., Schneider, M. A., Schrotz-King, P., Tournigand, C., Tran Van Nhieu, J., Yamada, T., Zimmermann, J., Benes, V., Kloor, M., Ulrich, C. M., von Knebel Doeberitz, M., Sobhani, I., & Bork, P.

(2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol*, *10*(11), 766. https://doi.org/10.15252/msb.20145645

Zellner, G., Alten, C., Stackebrandt, E., Conway de Macario, E., & Winter, J. (1987). Isolation and characterization of Methanocorpusculum parvum, gen. nov., spec. nov., a new tungsten requiring, coccoid methanogen. *Archives of Microbiology*, *147*(1), 13-20. https://doi.org/10.1007/BF00492898

Zellner, G., Stackebrandt, E., Messner, P., Tindall, B. J., Conway de Macario, E., Kneifel, H., Sleytr, U. B., & Winter, J. (1989). Methanocorpusculaceae fam. nov., represented by Methanocorpusculum parvum, Methanocorpusculum sinense spec. nov. and Methanocorpusculum bavaricum spec. nov. *Arch Microbiol*, *151*(5), 381-390. https://doi.org/10.1007/bf00416595

Zellner, G., & Winter, J. (1987). Secondary alcohols as hydrogen donors for CO2-reduction by methanogens*. *Fems Microbiology Letters*, *44*(3), 323-328. https://doi.org/https://doi.org/10.1111/j.1574-6968.1987.tb02309.x

Zhang, C., Xu, B., Lu, T., & Huang, Z. (2019). Metagenomic Analysis of the Fecal Microbiomes of Wild Asian Elephants Reveals Microflora and Enzymes that Mainly Digest Hemicellulose. *J Microbiol Biotechnol*, *29*(8), 1255-1265. https://doi.org/10.4014/jmb.1904.04033

Zhang, H., DiBaise, J. K., Zuccolo, A., Kudrna, D., Braidotti, M., Yu, Y., Parameswaran, P., Crowell, M. D., Wing, R., Rittmann, B. E., & Krajmalnik-Brown, R. (2009). Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci U S A*, *106*(7), 2365-2370. https://doi.org/10.1073/pnas.0812600106

Zhang, Y., & Gladyshev, V. N. (2008). Molybdoproteomes and evolution of molybdenum utilization. *J Mol Biol*, *379*(4), 881-899. https://doi.org/10.1016/j.jmb.2008.03.051

Zhao, G., Liu, J., Liu, X., Chen, M., Zhang, H., & Wang, P. G. (2007). Cloning and characterization of GDP-perosamine synthetase (Per) from Escherichia coli O157:H7 and synthesis of GDP-perosamine in vitro. *Biochemical and Biophysical Research Communications*, *363*(3), 525-530. https://doi.org/https://doi.org/10.1016/j.bbrc.2007.08.184

Zhao, Y., Boone, D. R., Mah, R. A., Boone, J. E., & Xun, L. (1989). Isolation and Characterization of Methanocorpusculum labreanum sp. nov. from the LaBrea Tar Pits. *International Journal of Systematic and Evolutionary Microbiology*, *39*(1), 10-13. https://doi.org/https://doi.org/10.1099/00207713-39-1-10

Zhong, Q., Kobe, B., & Kappler, U. (2020). Molybdenum Enzymes and How They Support Virulence in Pathogenic Bacteria. *Front Microbiol*, *11*, 615860. https://doi.org/10.3389/fmicb.2020.615860

Zhou, C., Zhao, H., Xiao, X. Y., Chen, B. D., Guo, R. J., Wang, Q., Chen, H., Zhao, L. D., Zhang, C. C., Jiao, Y. H., Ju, Y. M., Yang, H. X., Fei, Y. Y., Wang, L., Shen, M., Li, H., Wang, X. H., Lu, X., Yang, B., Liu, J. J., Li, J., Peng, L. Y., Zheng, W. J., Zhang, C. Y., Zhou, J. X., Wu, Q. J., Yang, Y. J., Su, J. M., Shi, Q., Wu, D., Zhang, W., Zhang, F. C., Jia, H. J., Liu, D. P., Jie, Z. Y., & Zhang, X. (2020). Metagenomic profiling of the pro-inflammatory gut microbiota in ankylosing spondylitis. *J Autoimmun*, *107*, 102360. https://doi.org/10.1016/j.jaut.2019.102360

Zhu, C., Zhang, J., Tang, Y., Zhengkai, X., & Song, R. (2011). Diversity of methanogenic archaea in a biogas reactor fed with swine feces as the mono-substrate by mcrA analysis. *Microbiol Res*, *166*(1), 27-35. https://doi.org/10.1016/j.micres.2010.01.004