Australian Government
Department of Agriculture

CSIRO

mla
MEAT & LIVESTOCK AUSTRALIA

final report

Project code: B.CCH.6610

Prepared by: Carly Rosewarne

Commonwealth Scientific and Industrial Research Organisation (CSIRO)

Date published: August 2015

ISBN: 9781741919448

PUBLISHED BY
Meat & Livestock Australia Limited
Locked Bag 991
NORTH SYDNEY NSW 2059

Metagenomic approaches for understanding the functional metabolic potential of methanogen communities in ruminant livestock

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication.

Acknowledgements

This project is supported by funding from Meat & Livestock Australia; and the Australian Government Department of Agriculture as part of the Carbon Farming Futures Filling the Research Gap Program.

We acknowledge collaborations from within the National Livestock Methane Program, including rumen fluid samples obtained from projects 01200.029/B.CCH.6420 and 01200.035/B.CCH.6510. Rumen fluid samples were also obtained from Xixi Li and Phil Vercoe (University of Western Australia); and Peter Kennedy and Ed Charmley (CSIRO)

Executive summary

Many surveys continue to be undertaken with the aim of cataloguing the key microbes responsible for methane production in livestock ("who's there?"), but these studies have been unable to provide functional information to describe their behaviour in the rumen ("what are they doing?"). There is a paucity of knowledge about how the genetic potential of methanogens is expressed to support their growth and (or) methane producing activity. By recovering, examining and comparing the genomes and gene expression profiles of rumen methanogens from Australian production systems, we have been able to reveal mechanisms that are critical for their persistence *in vivo*. This information has filled a significant research gap and can now be translated into a precise focus (rather than an empirical "shotgun" approach) for development of targeted approaches for inhibition of methanogenesis in ruminants.

Table of Contents

1. Background

It has been estimated that methane producing microbes (methanogens) only represent a small fraction of the total microbial population in the rumen (0.3-3.3% of small subunit rRNA genes; Janssen and Kirs, 2008). Techniques for determining which species are present ("who's there?") have consistently shown that there are three main types of methanogens in ruminants: autotrophs that use hydrogen to reduce carbon dioxide to methane and water (dominated by members of the *Methanobrevibacter* genus); methylotrophs that use hydrogen to reduce methanol to methane and water (dominated by members of the *Methanosphaera* genus); and methylotrophs that produce methane from methanol and/or methylated amines using a novel (coenzyme F_{420}-independent) pathway (dominated by the recently described 7th order of methanogens known as the Methanomassiliicoccales). Acetoclastic methanogens (those that use acetate to make methane) have been cultured from rumen samples but do not comprise a significant proportion of the community (Janssen and Kirs, 2008). An overview of the electron transport reactions used by the predominant methanogens in the rumen is provided in Figure 1.

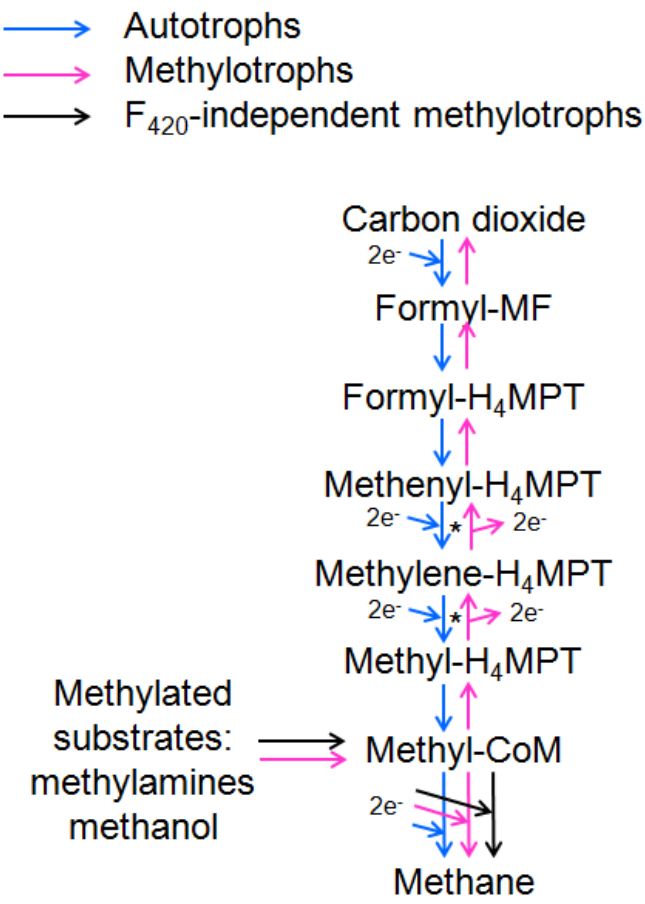


Figure 1. Simplified (unbalanced) schematic of the predominant methanogenesis pathways used by rumen methanogens. Blue arrows represent the autotrophic pathway; pink arrows represent the F_{420}-dependent methylotrophic pathway and black arrows represent the F_{420}-independent methylotrophic pathway. The reactions that can use coenzyme F_{420} cofactors for electron transport (methenyl-H_4MPT to methylene-H_4MPT and methylene-H_4MPT to methyl-H_4MPT) are indicated with an asterisk

Much of the literature that focuses specifically on methanogens in ruminants is descriptive in nature (e.g. determining differences in species composition due to diet or host) or is focused on analysis of individual strains isolated in pure culture. Methanogens are obligate anaerobes that grow slowly *in vitro* and often behave differently when cultured in the absence of other microbes (Samuel *et al*., 2007; Leahy *et al*., 2010). Furthermore, strains that are amenable to laboratory culture may not necessarily reflect those that predominate in the rumen. In this project, we set out to generate an in-depth understanding of methanogens relevant to methane production in Australian livestock production systems using novel techniques that would allow us to overcome the limitations inherent to these more traditional approaches.

2. Methods

Transcriptomic response of *Methanobrevibacter ruminantium* M1 to increasing concentrations of selected metals *in vitro*

- Growth medium for this experiment is described in detail in Appendix 1. The media was identical for all conditions except for the volume of metal solution 2 (provides exogenous $FeSO_4$, $CoSO_4$, $ZnSO_4$, Na_2MoO_4, $NiSO_4$, Na_2SeO_4 and Na_2WO_2). These metals are important for the function of many enzymes in the central methanogenesis pathway (Kaster *et al*., 2011)

- The concentration of iron, cobalt, zinc, molybdenum, nickel, selenium and tungsten in the growth media and rumen fluid stock was determined by the National Measurement Institute (NMI) using HR ICP-MS (high resolution inductively coupled plasma mass spectrometry)

- *M. ruminantium M1* cells were allowed to adapt to “low”, “medium” and “high” metal media for two passages before being inoculated into quadruplicate 50 mL cultures (1:50 dilution) in 120 mL bottles. The OD_{595} of each starter culture was 0.1. Bottles were gassed with 170 kPa of 4:1 H_2:CO_2 gas mix and incubated at 39 °C in the dark with gentle shaking (50 rpm)

- Cells were harvested at an OD595 of 0.1 (120 hours post-inoculation) in 10 mL of a stop solution (1:10 phenol:ethanol) to inhibit transcription during a short incubation on ice. Cells were recovered by centrifugation, resuspended in 1 mL of RNAlater (Ambion) and incubated on wet ice for 1 hour. Cells were recovered by centrifugation before the supernatant was discarded and the pellet stored at -80 °C until use

- RNA was extracted and purified using the PowerMicrobiome and RTS DNase removal kits (MoBio). RNA and DNA were quantified using the Qubit fluorimeter. Ribosomal RNA transcripts were removed using the RiboZero Magnetic kit for Bacteria (Epicenter). RNA quality pre and post-treatment was assessed using the Bioanalyzer RNA Pico kit (Agilent)

- Samples were shipped to Macrogen Inc. (Korea) as ethanol precipitates on dry ice, where sequencing libraries were prepared from ribo-depleted RNA samples using the Illumina TruSeq RNA preparation kit V2. Libraries were pooled and sequenced on a lane of Illumina HiSeq 2000 (2 x 101 bp paired-end reads)

- Raw sequencing data was filtered for downstream analysis using a custom wrapper script. Read quality pre and post filtering was assessed with FastQC v0.10.1. Paired reads were mapped against the reference genome (Leahy *et al*., 2010) using BWA-MEM v0.7.5a with default settings (Li, 2013). Resultant SAM files was compressed into a BAM format using SAMtools v0.1.19. The final quality of the alignment was manually assessed by visualisation in Geneious (v6.1.4; http://www.geneious.com/)

- Differential expression of counted reads was assessed using the negative binomial approach implemented in the R library DESeq2 (Anders and Huber., 2010). The betaPrior function was set to false as suggested by the authors. Differentially expressed genes were defined at p-value cutoff of 0.01

Amplicon profiling of methanogen population structure in Australian production systems

- Frozen rumen fluid samples (10 mL) were provided by Peter Kennedy and Ed Charmley, CSIRO (Brahman steers consuming a range of tropical grasses and legumes; refer to Kennedy and Charmley, 2012) and Xixi Li and Phil Vercoe, University of Western Australia (Merino x Suffolk wethers consuming an oaten chaff diet +/- supplementation with *Eremophila glabra;* unpublished data)

- DNA was extracted from 0.5 mL aliquots of rumen fluid using the repeated bead-beating and column (RBB+C) method (Yu and Morrison, 2004), optimised for the PowerSoil DNA extraction kit (MoBio)

- Primers specific for the archaeal 16S ribosomal RNA gene (A340F and A1000R; Gantner *et al*., 2011) were used to amplify PCR products from 10 ng of each DNA sample in duplicate, using the protocol described by Aguirre *et al*., 2011a. Note that the number of PCR cycles in the first round was increased from 20 to 25

- Uparse (Edgar, 2013) was used for assigning sequences to barcodes, read quality filtering and length trimming (400 base pairs per sequence), dereplication, chimera checking and clustering of sequences into operational taxonomic units (OTUs; groups of related sequences) at a sequence similarity cutoff of 97% (approximately equivalent to a species-level grouping). The resulting

OTU table was filtered to remove low abundance clusters (OTUs that contain less than 5 sequences or cannot be found in at least three samples)

- The representative sequence from each OTU was classified using the Ribosomal Database Project taxonomic hierarchy (Cole *et al*., 2011) - any OTUs not recognised as archaeal in origin were also filtered from the OTU table. The table was then rarefied to the median number of sequences per sample in the dataset (justification as described by Aguirre *et al*., 2011b). In a recent study it was demonstrated that 1,000 sequences is sufficient to generate an accurate survey of rumen methanogen community composition (Kittlemann *et al*., 2013). For our samples, the number of sequences ranges from 826 to 1,504

- Statistically significant differences between the relative abundance of key taxa were determined using multiple t tests in GraphPad Prism 6.0 (Holm-Sidak method, α=5%, without assumption of a consistent standard deviation)

- All OTUs from the *Methanobrevibacter* genus that contained more than 1% of the total sequences from either dataset were identified (n=17). Collectively, these OTUs contain 86% of the total data from cattle and 93% of the total data from sheep. A representative sequence from each OTU, and 16S rRNA gene sequences from *Methanobrevibacter* strains were aligned using PyNAST in QIIME before constructing a phylogenetic tree with phylogeny.fr (Dereeper *et al*., 2008)

Production of a comprehensive metagenomic database of rumen methanogen functional potential in Australian production systems

- Rumen fluid from 3 Brahman steers consuming Flinders grass hay (*Iseilema sp*.) was provided by Nigel Tomkins, CSIRO (as part of project 01200.029/B.CCH.6420). Samples were collected through a rumen cannula and filtered through 150 μm nylon mesh before freezing on dry ice in 25 mL aliquots

- A wet laboratory protocol was developed to increase the relative abundance of *Methanobrevibacter* DNA in metagenomic DNA extracted from rumen fluid. Details may be requested by emailing the primary contact for this project. The success of different treatments was quantified using real-time PCR with primers targeting the methanogen mcrA gene (Denman *et al*, 2007) and the bacterial 16S rRNA gene (Denman *et al*., 2006)

- DNA from rumen samples treated to increase the relative abundance of *Methanobrevibacter* DNA were sequenced using Illumina HiSeq 2000 technology (Macrogen Inc., Korea; 2 x 101 bp reads

from TruSeq Nano 550 bp insert libraries). Approximately 7.4 x 10^7 reads were obtained for each sample (n=3). A suite of bioinformatics tools were used to analyse the data:

- Sequence read quality trimming - Trimmomatic (Bolger *et al*., 2014)
- Identification of ribosomal RNA gene fragments - RNammer (Lagesen *et al*., 2007)
- Classification of 16S rRNA gene fragments - Ribosomal Database Project (Cole *et al*., 2014)
- Comparison of metagenome 16S rRNA gene fragments to type strains - closed reference OTU picking in QIIME at 97% similarity cutoff (Caporaso *et al*., 2010)
- Metagenome assembly - IDBA-UD (Peng *et al*., 2012) using --pre_correction and --mink=40 flags
- Taxonomic "binning" of large (>1 kb) metagenomic contigs - PhyloPythiaS web server (Patil *et al*., 2012). The "Generic 2013 - 800 genera" in-built model was used to identify contigs that are archaeal in origin
- Gene prediction - MetaGeneMark (Zhu *et al*, 2010)
- Annotation - Blast2GO (Gotz *et al*, 2008) and InterProScan 5 (Jones *et al*., 2014)
- Identification of cell-surface exposed proteins - Philius (Reynolds *et al*., 2008)

Characterisation of rumen methanogen gene expression in northern Australian beef cattle

- Rumen fluid from 10 steers consuming improved Rhodes grass pasture (*Chloris gayana*) +/- supplementation with *Leucaena* was provided by Chris McSweeney, CSIRO (as part of project 01200.035/B.CCH.6510). Samples were collected via stomach tube and filtered through a metal sieve (to remove large particulates) before freezing on dry ice in 5 mL aliquots

- For each sample, an methanogen metagenomic database was prepared as described in the preceding section

- Total RNA was extracted from rumen fluid using a modified version of the protocol described by Piao *et al* (2013). Briefly, aliquots of rumen fluid were thawed on wet ice and strained through 150 μm nylon mesh. Duplicate 500 μL aliquots were removed and centrifuged at 10,000 x g for 5 minutes at 4 °C to pellet cells. Cells were resuspended in 1 mL Trizol reagent (Life Technologies) and transferred to 1.5 mL screw-cap tubes containing 0.1 mm sterile glass beads. Samples were homogenised for 1 minute at maximum speed (FastPrep24, MP Biomedicals) followed by a 2 minute incubation on ice x 3. Samples were centrifuged at 10,000 x g for 10 minutes at 4 °C before being purified using the Direct-zol RNA miniprep kit (Zymo Research), including the optional step for on-column DNase treatment

- Duplicate RNA samples were pooled and processed sequentially using the RiboZero Magnetic Gold kit for Epidemiology, followed by the RiboZero Magnetic Gold kit for Bacteria (Epicenter). A minimum of 100 ng of ribo-depleted RNA for each animal was sequenced using Illumina HiSeq 2000 technology (2 x 101 bp reads from TruSeq RNA V2 libraries). Approximately 8 x 10^7 reads were obtained for each library (n=9; one sample failed to pass QC due to low input RNA concentration upon receipt at Macrogen). A suite of bioinformatic tools was used to analyse the data:

    - Sequence read quality trimming - Trimmomatic (Bolger *et al*., 2014)
    - Identification of ribosomal RNA gene fragments - RNammer (Lagesen *et al*., 2007)
    - Identification of methyl coenzyme reductase A (mcrA) transcripts - hmmer (Eddy, 2011) using hidden Markov models from Pfam (Finn *et al*., 2014) that are specific to mcrA (MCR_alphaN: PF02745; and MCR_alpha: PF02249)
    - Comparison of metatranscriptome mcrA transcripts to full length mcrA genes - closed reference (uclust_ref) OTU picking in QIIME at 92% similarity cutoff (Caporaso *et al*., 2010), using mcrA genes from the methanogen metagenome database, from methanogen genomes deposited in the IMG database (http://img.jgi.doe.gov) and the study published by Shi *et al* (2014)
    - Alignment of mRNA reads against methanogen genes - Bowtie 2 (Langmead and Salzberg, 2012)
    - Identification of highly expressed genes - HTseq (Anders *et al*., 2015)
    - Annotation of highly expressed genes - Blast2GO (Gotz *et al*, 2008) and InterProScan 5 (Jones *et al*., 2014)

3. Results

Transcriptomic response of *Methanobrevibacter ruminantium* M1 to increasing concentrations of selected metals *in vitro*

Many enzymes involved in the central methanogenesis pathway contain metal cofactors that are essential for their function. Methanogens need to be able to facilitate uptake of these metals from their environment, co-ordinated by specific ion transporters. These transporters are located on the cell surface and as such they represent good targets for developing novel methane abatement strategies. Under metal limiting conditions the methanogens will need to work harder in order to extract transition metals from their environment. We hypothesised that *M. ruminantium* M1 (the best characterised rumen methanogen strain) would overexpress the genes required for metal uptake when grown in metal-depleted media, and that a comparative transcriptomics approach may allow us to identify them.

The first step in this experiment involved optimising the growth of M1 in a metal limited broth. The standard medium (as used in our laboratory) contains 10% clarified rumen fluid and 2g/L of yeast extract and tryptone. All of these components support the growth of M1 but also deliver trace concentrations of metals. Broth was prepared to contain 0.1x, 0.25x and 1x the amount of rumen fluid, yeast extract and tryptone found in the standard recipe. The volume of metal solution 2 (iron, cobalt, zinc, molybdenum, nickel, selenium and tungsten) was also adjusted to reflect 0x, 1x and 10x the amount in the standard recipe. Growth curves (OD_{595}) were monitored over a 168 hour period (Figure 2). This showed that M1 is incapable of growth in any broth with 0.1x rumen fluid, yeast extract and tryptone. Growth rates in all cultures with 0.25x rumen fluid, yeast extract and tryptone were slightly decreased when compared to the 1x broth, irrespective of the amount of metal solution that was added. However, since there was no appreciable difference in growth rate in the 0.25x broth derivations containing 0x, 1x or 10x metal solution, these preparations were chosen for further analysis.

The concentration of metals in each derivation of broth containing 2.5% rumen fluid and 0.5 g/L of yeast extract and tryptone (hereafter referred to as “low”, “medium” and “high” to reflect the addition of metal solution) was determined using HR ICP-MS (Table 1). The concentrations of all metals except zinc are below the limit of detection for the low metal broth. There is at least a 10-fold increase in the concentration of cobalt, molybdenum, nickel, selenium and tungsten between the low and medium broths and the medium and high broths; and a 5-fold increase in the concentration of zinc. Iron is below the detection limit for the low and medium broths, and there is a 3-fold increase between the medium and high broths. Physiological trace metal concentrations in the rumen, based on comparison to the rumen fluid used in broth preparation, are between the low and medium broths for all elements except iron, selenium and zinc.

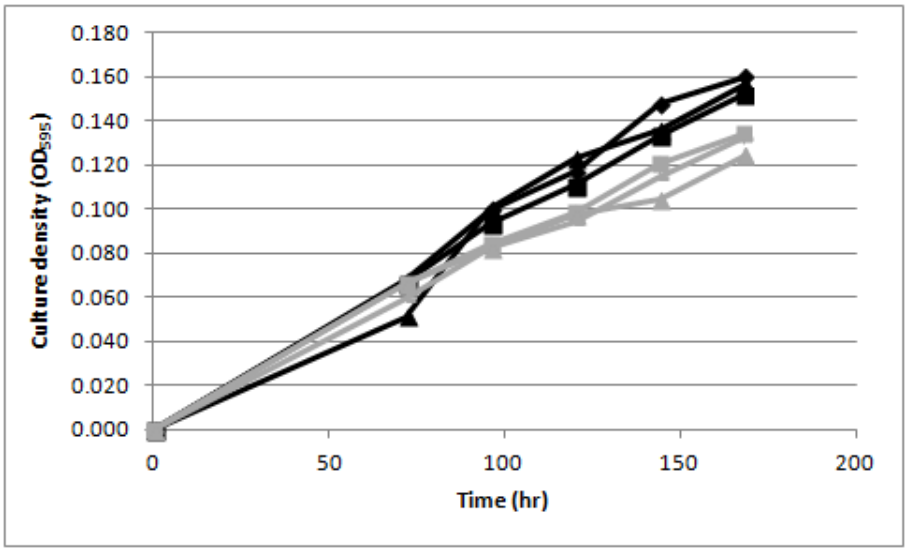


Figure 2. Growth curves of *Methanobrevibacter ruminantium* M1 in broth containing variable concentrations of rumen fluid, yeast extract and tryptone (0.25x standard: grey curves; 1x standard: black curves) and metal solution (0x standard: diamond; 1x standard: square; 10x standard: triangle)

Table 1. Determination of trace metal concentrations in broth preparations and rumen fluid using HR ICP-MS. Results are reported in parts per million. The lower limit of detection for each element is indicated in parentheses

Trace Element	Broth with no metal solution ("low")	Broth with 1x metal solution ("medium")	Broth with 10x metal solution ("high")	Rumen fluid
Cobalt (0.01)	<0.01	0.23	2.3	0.05
Iron (0.5)	<0.5	<0.5	1.4	1
Molybdenum (0.01)	<0.01	0.30	2.7	0.01
Nickel (0.01)	<0.01	0.13	1.2	0.09
Selenium (0.05)	<0.05	0.45	4.6	<0.05
Tungsten (0.01)	<0.01	0.35	3.5	0.04
Zinc (0.01)	0.11	0.48	2.5	0.77

Based on the HR ICP-MS results, the fold differences in concentration between the three broth derivations would be sufficient to detect changes in gene expression using transcriptomics. M1 cells were allowed to adapt to each broth before being subcultured into quadruplicate 50 mL cultures. Cultures were harvested after 120 hours of incubation, when the OD_{595} had reached 0.1. Messenger RNA (mRNA; n=12) was sequenced in one lane using Illumina HiSeq 2000 (100bp paired-end reads). An average of 1 x 10^7 reads per sample were included in the final analysis. Figure 3 shows a heatmap of the top 50 differentially expressed genes (at an adjusted p value <0.01). Cells grown in the low metal broth are significantly different to those grown in medium or high metal broths, while cells in the medium and high metal broths are similar. As a result the downstream analysis was focused on comparing the profiles between the low and medium metal broth samples using DEseq. A total of 673 out of 2217 genes in the M1 genome were considered to be differentially expressed between these two conditions.

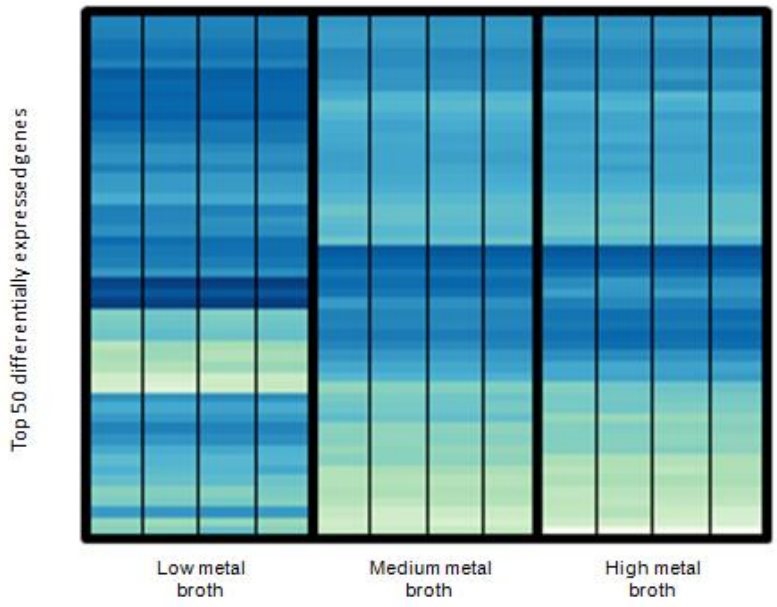


Figure 3. Heatmap depicting the top 50 differentially expressed genes in *M. ruminantium* M1 grown in broth cultures with low, medium and high concentrations of trace elements.

Using the DEseq analysis we were able to identify genes relevant to metal acquisition that were upregulated under metal limiting conditions. The four genes exhibiting the highest differential expression are found in an operon that encompasses genes mru_0251 to mru_0254. Although expression occurs under both conditions, it is increased by more than 20-fold due to metal limitation (Figure 4). Genes mru_0251, 252 and 253 encode proteins with homology to ATP binding cassette (ABC) transporters. These three-component systems use ATP to drive translocation of substrates across cell membranes. The specific substrate(s) transported by this system in M1 are not known (Leahy *et al*., 2010). Genes for transport of ferrous iron (FeoA/B2; genes mru_0536-7) and binding/translocation of heavy metals using a P-type ATPase system (genes mru_0205-6) were also upregulated. Interestingly, there are two operons annotated as nickel ABC transporters in the M1 genome, they are denoted as Nik1 (genes mru_1614-8) and Nik2 (genes mru_1705-10). In the metal limited culture, expression of Nik1 is increased but Nik2 is not. This may suggest that different mechanisms operate depending on substrate availability.

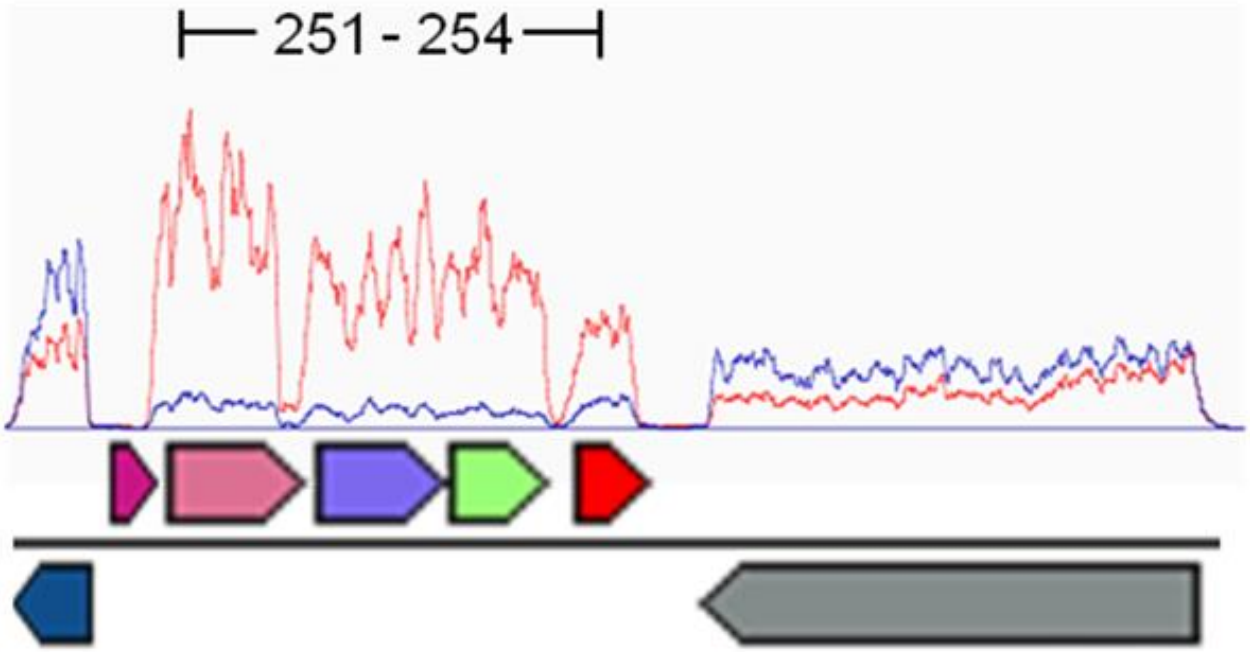


Figure 4. Diagrammatic representation of gene expression from genes mru_0249 to mru_0255 in *M. ruminantium* M1 grown in low (blue) and medium (red) metal broths

The aim of this experiment was to identify systems used by methanogens for metal transport under limiting conditions. Since many metal ions are essential for the function of enzymes in the central methanogenesis pathway, inhibition of specific uptake mechanisms may provide a pathway for development of targeted methane mitigation strategies. For such strategies to be effective, targets must be conserved in rumen methanogens but distinct from those found in rumen bacteria. Searches within the Integrated Microbial Genomes database (IMG; http://img.jgi.doe.gov) revealed that proteins encoded

by each of the upregulated genes share homology with proteins found in members of the *Bacteroidetes* and *Firmicutes*, the two most common bacterial phyla in vertebrate gastrointestinal microbiomes (Ley *et al*., 2008). Based on this result, we did not conduct any further analysis on these upregulated genes.

Amplicon profiling of methanogen population structure in Australian production systems

In this analysis, we set out to produce a detailed survey of methanogen community structure ("who's there?') in two cohorts of ruminants representative of Australian production systems. This was achieved using a molecular profiling approach, which allows identification of the methanogens present in rumen fluid based on determination of the DNA sequence composition of a single marker gene (16S ribosomal RNA gene).

The first set of rumen fluid samples was provided by Peter Kennedy and Ed Charmley from CSIRO. They were collected as part of a published study investigating methane yields from Brahman cattle fed tropical grasses and legumes (Kennedy and Charmley, 2012). A description of the dataset, which encompasses a total of 16 different diets (80 samples tested in 13 individual animals) is provided in Appendix 2.

The methanogen population in this cohort of northern Australian beef cattle is dominated by the genus *Methanobrevibacter* (Figure 5). Members of this genus comprise between 76-98% of the methanogen population depending on the forage. Feed comprised of high quality Rhodes grass +/- supplementation with 20-40% *Stylosanthes hamata* cv. Verano or 20-40% *Macroptilium bracteatum* (Burgundy bean) resulted in the highest proportion of methanogens belonging to the Methanomassilicoccales (between 5-11%).

The second set of rumen fluid samples was provided by Xixi Li and Phil Vercoe from the University of Western Australia. As part of her PhD, Xixi was investigating the antimethanogenic properties of *Eremophila glabra* (emu bush) in sheep, finding that supplementation reduced methane emissions by 11.8% per gram of dry matter (DM) intake and 14.8% per gram of digested DM. Further information about this study can be found in Appendix 2.

The methanogen population in this cohort of sheep is also dominated by genus *Methanobrevibacter*, comprising 92-99% of sequences from each animal (Figure 6a). Differences in methane emissions between the two diets may be related to the decreased relative abundance of methanogens belonging to the Methanomassiliicoccales in the test samples (4.5 +/- 1.8 for the control samples vs 1.3 +/- 1.4 for the test samples), concomitant with a relative increase in the relative abundance of both the *Methanobrevibacter* and *Methanosphaera* genera (Figure 6b).

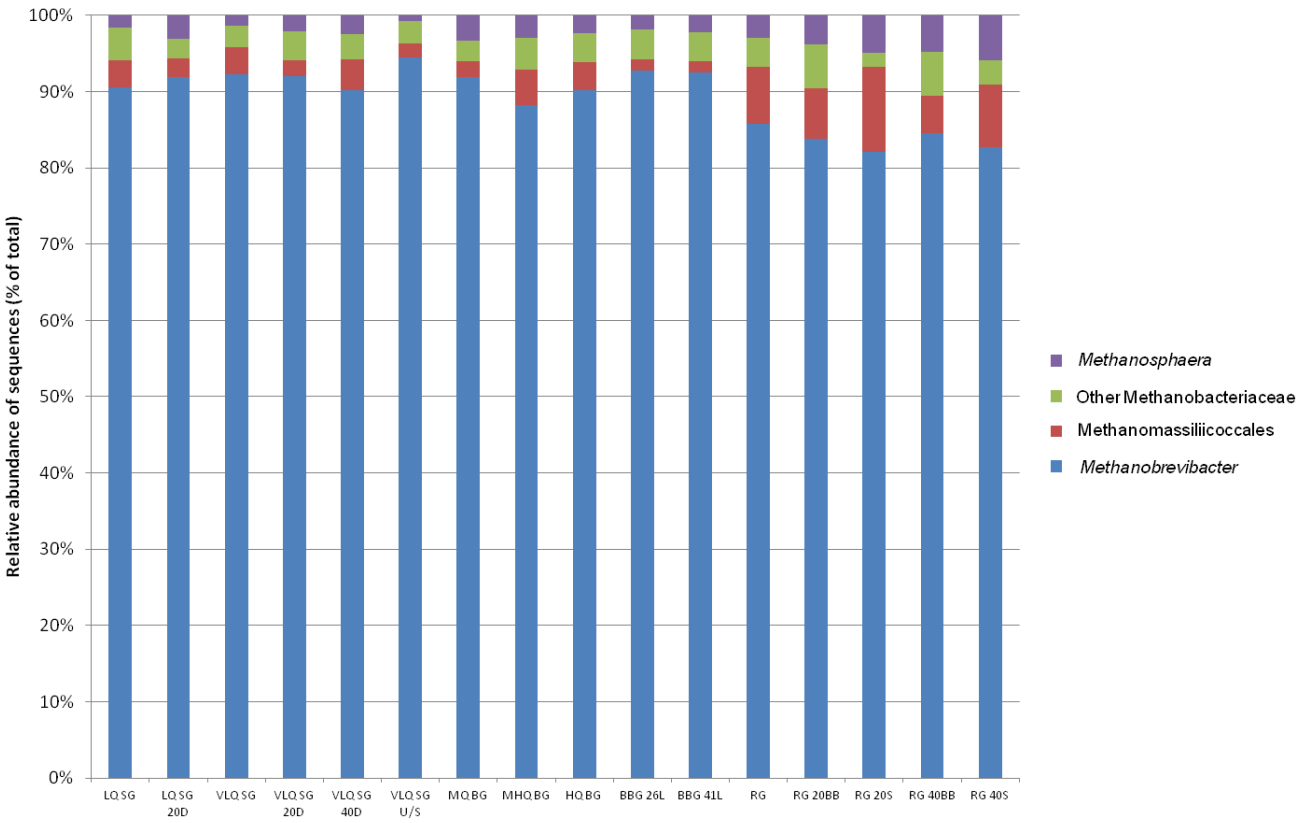


Figure 5. Rumen methanogen population structure in northern Australian beef cattle consuming tropical forages

- LQ SG: low quality black speargrass
- LQ SG 20D: low quality black speargrass + 20% *Dolichos*
- VLQ SG: very low quality speargrass
- VLQ SG 20D: very low quality black speargrass + 20% *Dolichos*
- VLQ SG 20D: very low quality black speargrass + 40% *Dolichos*
- VLQ SG U/S: very low quality black speargrass + urea/sulfur supplementation
- MQ BG: medium quality buffel grass
- MHQ BG: medium to high quality buffel grass
- HQ BG: high quality buffel grass
- BBG 26L: bisset creeping bluegrass, 26% leaf
- BBG 41L: bisset creeping bluegrass, 41% leaf
- RG: Rhodes grass
- RG 20BB: Rhodes grass + 20% burgundy bean
- RG 20S: Rhodes grass+ 20% *Stylosanthes*
- RG 40BB: Rhodes grass + 40% burgundy bean
- RG 40S: Rhodes grass + 40% *Stylosanthes*

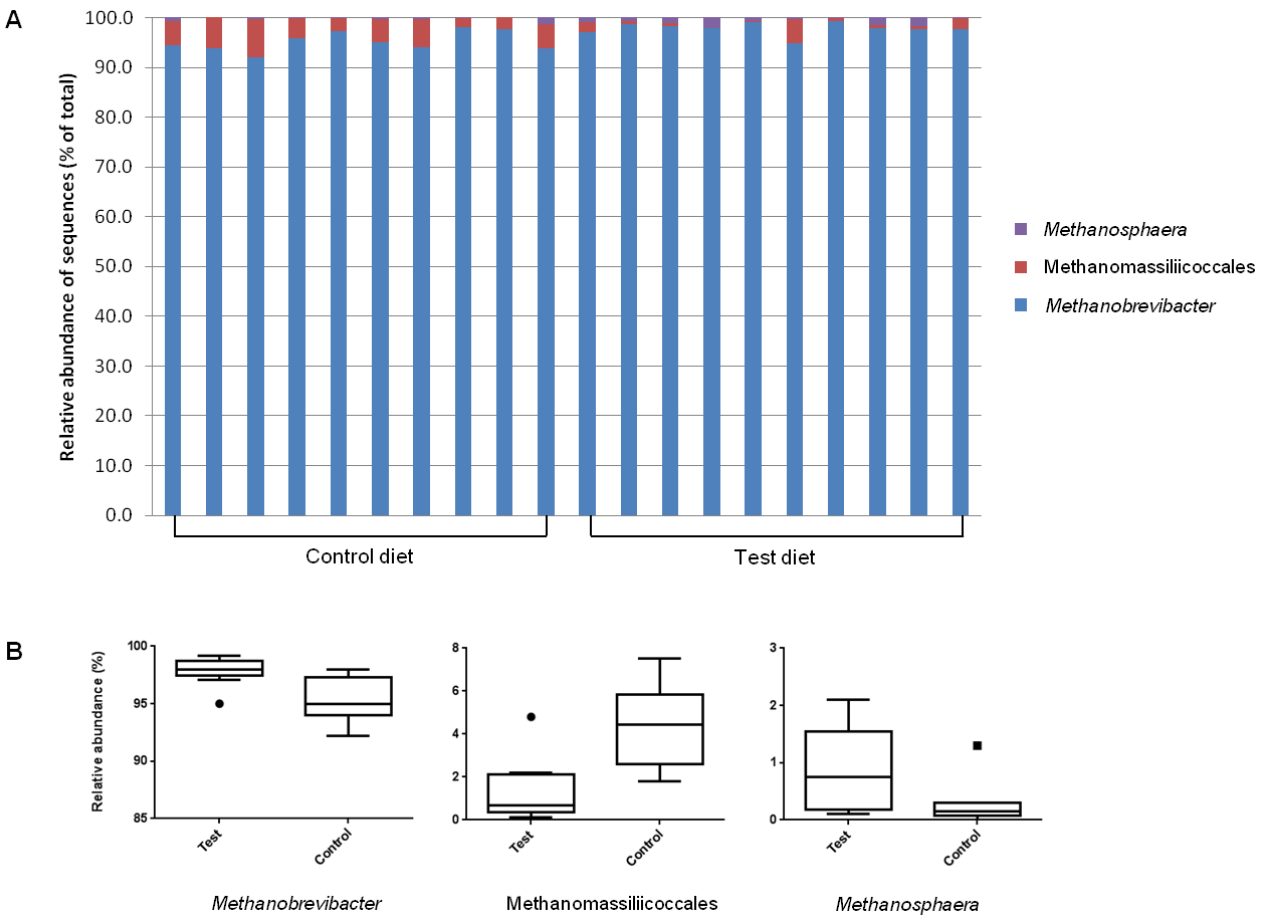


Figure 6. Rumen methanogen population structure in sheep consuming oaten chaff +/- 15% *Eremophila glabra*

A: Results from individual sheep (n=10 per treatment group)
B: Boxplot showing the relative abundance of *Methanobrevibacter,* Methanomassiliicoccales and *Methanosphaera* on each diet

Differences in relative abundance of each of the three groups of methanogens were significant between the test diet and control diet. Supplementation with *E. glabra* resulted in a lower relative abundance of *Methanobrevibacter* sp. ($p<0.002$) and *Methanosphaera* sp. ($p<0.05$) concomitant with a higher proportion of Methanomassiliicoccales ($p<0.0005$).

The results indicate that members of the genus *Methanobrevibacter* are the predominant rumen methanogens in both sample sets. Type strains from the genus *Methanobrevibacter* have been cultured from a range of samples including the digestive tract of ruminants, birds, monogastrics and termites; anaerobic waste fermenters and natural environments. The phylogenetic tree presented in Figure 7 demonstrates the relatedness of the cattle and sheep rumen *Methanobrevibacter* populations to each other and to the type strains based on comparison of 16S rRNA gene sequences. It shows that the *Methanobrevibacter* species composition is comparable in both the cattle and sheep cohorts, as similar numbers of sequences from both studies were attributed to the same phylogenetic groups (called OTUs, or operational taxonomic units).

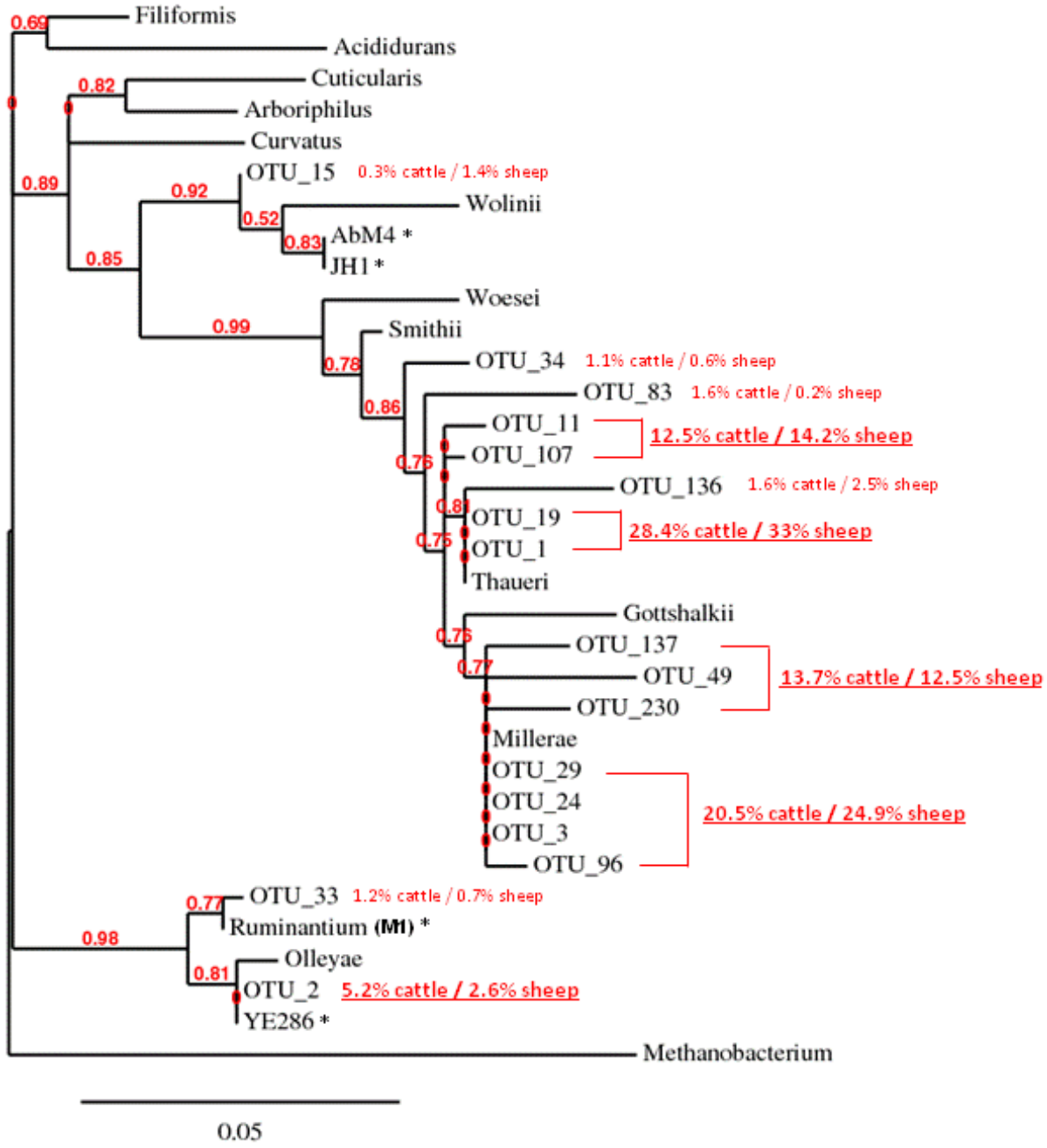


Figure 7. Phylogenetic tree of partial 16S rRNA gene sequences showing relationships between *Methanobrevibacter* isolates and OTUs identified during cattle and sheep population surveys. Predominant OTU clusters are underlined. Only OTUs containing more than 1% of the total sequences from either dataset are included on the tree. Strains for which genome sequence data was available at the time of analysis (n=4; produced for milestone report submitted 31/10/2013) are indicated with an asterisk. NCBI accession numbers for 16S rRNA genes from strains are listed in Leahy *et al*., 2013

Genome sequence information represents the first step in being able to link methanogen population structure ("who's there?") to function ("what are they doing?"). When Figure 7 was produced (Milestone report 3, submitted 31/10/2013) the *Methanobrevibacter* isolates for which genome sequence data was available (n=4) were not closely related to the species present in either the northern Australian beef cattle

or the southern Australian sheep rumen samples. Strain YE286 is related to OTU_2, which contains 5.2% of cattle sequences and 2.6% of sheep sequences. Strain M1 is related to OTU_33, which contains 1.2% of cattle sequences and 0.7% of sheep sequences, while strains AbM4 and JH1 are related to OTU_15, which contains 0.3% of cattle sequences and 1.4% of sheep sequences. A total of 75% of the total methanogens in cattle and 85% of the total methanogens in sheep were present in clusters with similarity to *M. thaueri*, *M. gottschalkii* and *M. millerae*. The next step in the project was to generate genome sequence data from the predominant *Methanobrevibacter* strains in Australian production systems.

Production of a comprehensive metagenomic database of rumen methanogen functional potential in Australian production systems

Interrogating genome sequences allows us to characterise the functional potential of rumen methanogens, and answer the question "what can they do?". All of the published information about rumen methanogen genomes has been generated from sequencing of isolated strains. The so-called "cultivation bias" means that the literature contains data from strains that grow well under *in vitro* (laboratory) conditions. These strains don't necessarily represent those that predominate under *in vivo* (rumen) conditions. A different approach was required in order to study the methanogen populations responsible for methane production in the rumen.

In a typical rumen fluid sample, methanogens comprise between 0.3-3.3% of the total microbial population. Standard shotgun metagenome sequencing approaches, used to analyse the DNA content of abundant microbes in a sample, would not be able to provide sufficient information for the study of methanogen genomes. We developed a protocol to increase the relative abundance of DNA from methanogens in rumen metagenomic samples. Three separate treatments were tested, both individually and in combination. When all three treatments were applied sequentially to the same test sample, the relative abundance of methanogen DNA increased from 1.5% to 18% (Figure 8).

The combination of three sequential treatments (T1 + T2 +T3) was applied to rumen fluid samples from three Brahman steers consuming Flinders grass hay. Samples were subjected to next-generation sequencing, which returned an average of 7.4×10^7 reads per sample (101 base pairs each). The ribosomal RNA reads were extracted using RNammer and 16S rRNA reads were classified using the Ribosomal Database Project (RDP) hierarchy. Analysis of the 16S rRNA reads gives a preliminary indication of the microbial composition of the enriched metagenomes and provides a more accurate measure of the proportion of methanogens relative to bacteria. For each sample, approximately 4×10^5 (0.5%) of the reads originated from a 16S rRNA gene. Classification indicated that 70-75% of the 16S rRNA reads from each sample were bacterial and 25-30% were archaeal in origin. All of the archaeal 16S rRNA reads from each sample were assigned to the Euryarchaeota. Approximately 85% of these reads were assigned to the *Methanobrevibacter* genus and 10% were assigned to the *Methanosphaera* genus.

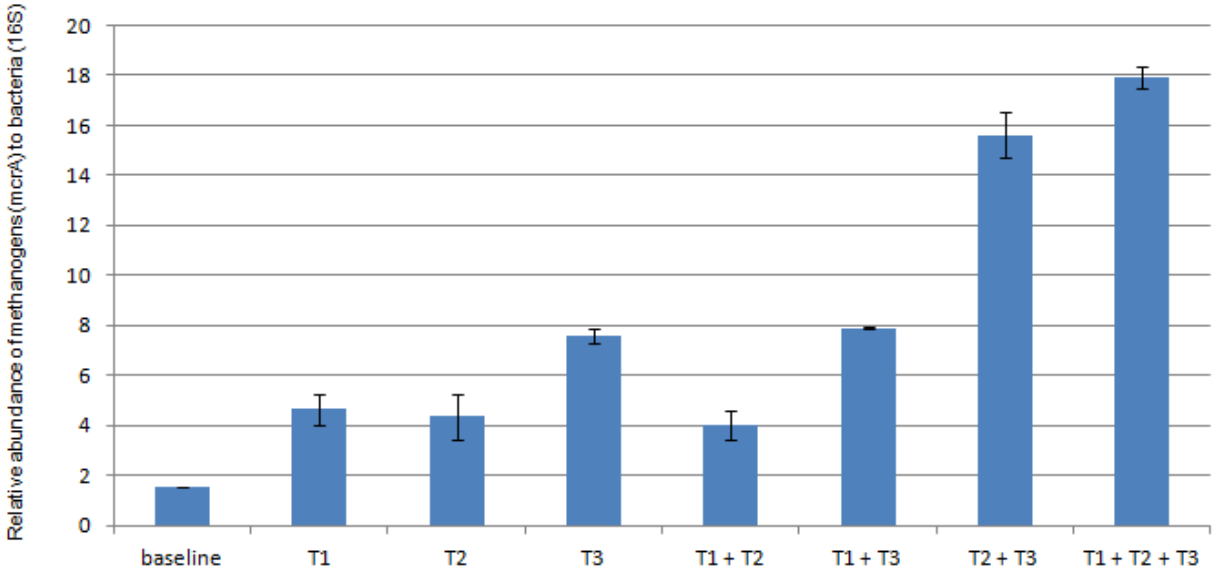


Figure 8. Relative percentage of methanogen mcrA to bacterial 16S rRNA genes in metagenomic DNA after treatments to increase the abundance of methanogens. Each treatment was tested in duplicate. Each duplicate treatment was quantified in triplicate. Error bars represent the standard error of six qPCR assays

Using the closed reference OTU clustering function in QIIME, the methanogen metagenomic 16S rRNA gene fragments were compared to near-full length 16S rRNA gene sequences from *Methanobrevibacter* type strains (97% similarity cutoff). Approximately 85% of the methanogen fragments were clustered with a type strain sequence, comparable to the estimate of *Methanobrevibacter* relative abundance. The *Methanobrevibacter* 16S rRNA gene fragments from the metagenomes are most similar to type strains *M. millerae*, *M. smithii* and *M. thaueri*, in addition to non-stype strain YE286 (Figure 9). Steer 434 has a high proportion of reads (19%) that are similar to non-type strain AbM4. Comparatively few reads cluster with sequences from *M. ruminantium* and *M. olleyae*, similar to the data presented in Figure 7. This suggests that the combination of three sequential treatments (T1 + T2 +T3) applied to these rumen samples prior to next generation sequencing has provided data that reflects the predominant species of *Methanobrevibacter* in Australian production systems.

All reads from the individual metagenomes were then assembled separately using IDBA-UD (Peng *et al*., 2012). This process creates contigs, which are contiguous stretches of genome sequence generated by overlapping the short read data. Longer contigs provide more information about the genetic potential of the organism from which they were derived. A summary of the output is shown in Table 2. Each assembled dataset contains around 45% of the total input reads, suggesting that a significant proportion of the data is encompassed by the large contigs (>1 Kb).

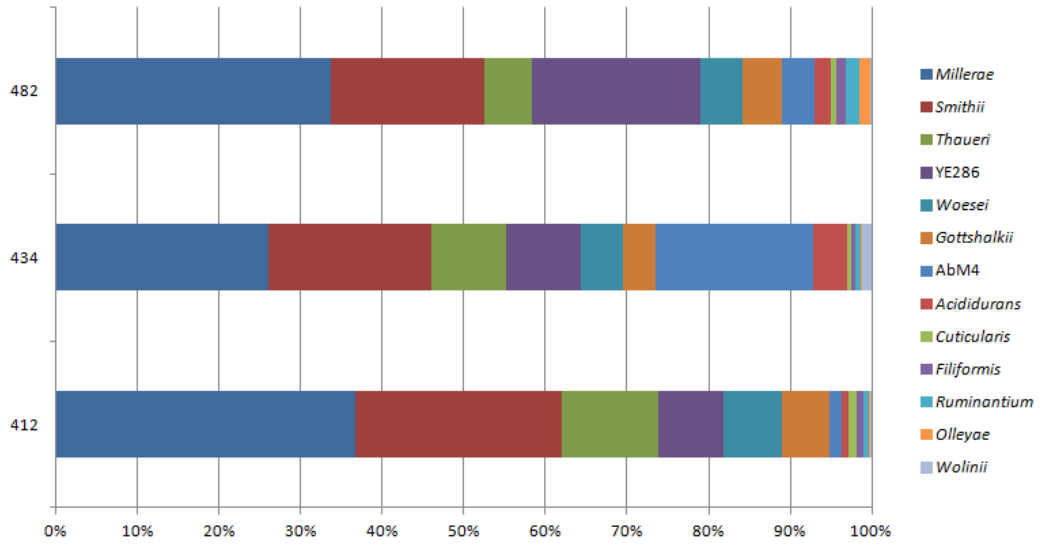


Figure 9. Comparison of 16S rRNA gene fragments from the pilot metagenome database to selected *Methanobrevibacter* strains

Table 2. Assembly data for each metagenome in the Flinders grass study (before separation of bacterial and archaeal contigs)

Sample (steer number)	Number of trimmed input reads	Number of large contigs (> 1Kb)	Size of assembled dataset (large contigs; Mb)	% of input reads used in assembly (large contigs)
412	6.51 x 10^7	96138	227.4	41.9
434	8.73 x 10^7	100980	234.5	45.3
482	6.90 x 10^7	71984	173.6	46.7

At this stage, the assembled dataset contains contigs from both methanogens and bacteria. We used the PhyloPythiaS online web server to classify large contigs from each assembly into domain level taxonomic "bins", thereby separating the archaeal (methanogen) contigs from the bacterial contigs (Table 3) and creating a methanogen metagenome from each animal. This process provided 3 datasets, each containing 32-46 Mb of data. The number of input reads aligned to archaeal contigs (17-23%) correlates with our estimate of the relative percentage of archaeal 16S rRNA genes in each metagenome (25-30%), suggesting that a large proportion of the sequence reads from methanogen genomes have been successfully assembled.

Table 3. Properties of enriched methanogen metagenome datasets from the Flinders grass study

Sample (steer number)	Number of methanogen contigs	Mean length/N50 (bp)	Total size (Mb)	% of input reads aligned to contigs	Total number of genes
412	8791	5483 / 9374	46.1	19.8	46550
434	6204	5397 / 9949	32.0	16.9	33053
482	8494	5514 / 8510	44.8	22.8	44975

Since the average size for a rumen *Methanobrevibacter sp*. genome is 2-3 Mb, each dataset potentially encompasses 10-20 individual strains. It is likely that many of the strains are present in all samples.

Given that there are only a handful of rumen methanogen genomes that are available in public repositories, the information contained within this database represents a significant step forward in our current understanding of the functional potential of these microbes. We identified genes on each contig using Metagenemark (*Zhu et al*., 2010), and used this information to assign functions to proteins in Blast2GO (Gotz *et al*., 2008). Blast2GO implements several methods for annotation, including BLAST and InterProScan 5. Assignment of a protein sequence to an InterPro family is based on recognition of defined domains and important sites, providing consistent functional predictions that can be used for comparative genomics.

Interrogation of the pilot rumen methanogen metagenome database reveals differences in substrate utilisation potential within the *Methanobrevibacter* genus

Members of the *Methanobrevibacter* genus use hydrogen to reduce carbon dioxide to methane. Some members of this genus express the formate dehydrogenase enzyme, facilitating the use of formate as an alternative electron donor. The genes encoding this enzyme are found in the genome of all sequenced type strains of *Methanobrevibacter* (Figure 10; fdhCAB; coloured in yellow). In all available examples from isolates, the fdhCAB operon is located in the same genomic region as the formylmethanofuran dehydrogenase operon (Figure 10; fwdCABDGFH; coloured in blue). This enzyme catalyses the first step in conversion of carbon dioxide to methane. In the metagenome dataset, the genes encoding formate dehydrogenase and formylmethanofuran dehydrogenase are collocated on 482 contig A, but on 482 contig B the formate dehydrogenase genes are missing. Since the genomic architecture is conserved in all known examples from this genus, it is probable that the methanogen from which 482 contig B was derived cannot use formate as an electron donor.

The dominant methanogens in the human gut are related to *Methanobrevibacter smithii*. The majority of sequenced *M. smithii* strains (23/24; Samuel *et al*., 2007; Hansen *et al*., 2011) encode a methanol:coenzyme M methyltransferase system (MtaC, MtaB and MtaA), facilitating conversion of methanol into methyl-coenzyme M at the penultimate step of the methanogenesis pathway (Figure 1; Samuel *et al*, 2007). This system has not been found in rumen *Methanobrevibacter* type *strains* (*M. ruminantium* M1, Leahy *et al*., 2010; *M. millerae* and *M. olleyae,* unpublished Hungate 1000 data available at http://img.jgi.doe.gov) or non-type strains (JH1, Lee *et al*., 2013; AbM4, Leahy *et al*., 2013; and YE286, unpublished Beef CRC resource). It is present in the genome of the sheep faecal type strain *M. wolinii* SH (unpublished Hungate 1000 data available at http://img.jgi.doe.gov). A search of the rumen methanogen metagenome database revealed three contigs encoding Mta proteins. BLASTn searches of the NCBI database indicated that two of the contigs are most similar to *Methanobrevibacter smithii* PS, while the other is most closely related to *Methanosphaera stadtmanae* DSM 3091, a human gut methanogen that is restricted to growth on methanol and hydrogen as substrates (Fricke *et al*., 2006).

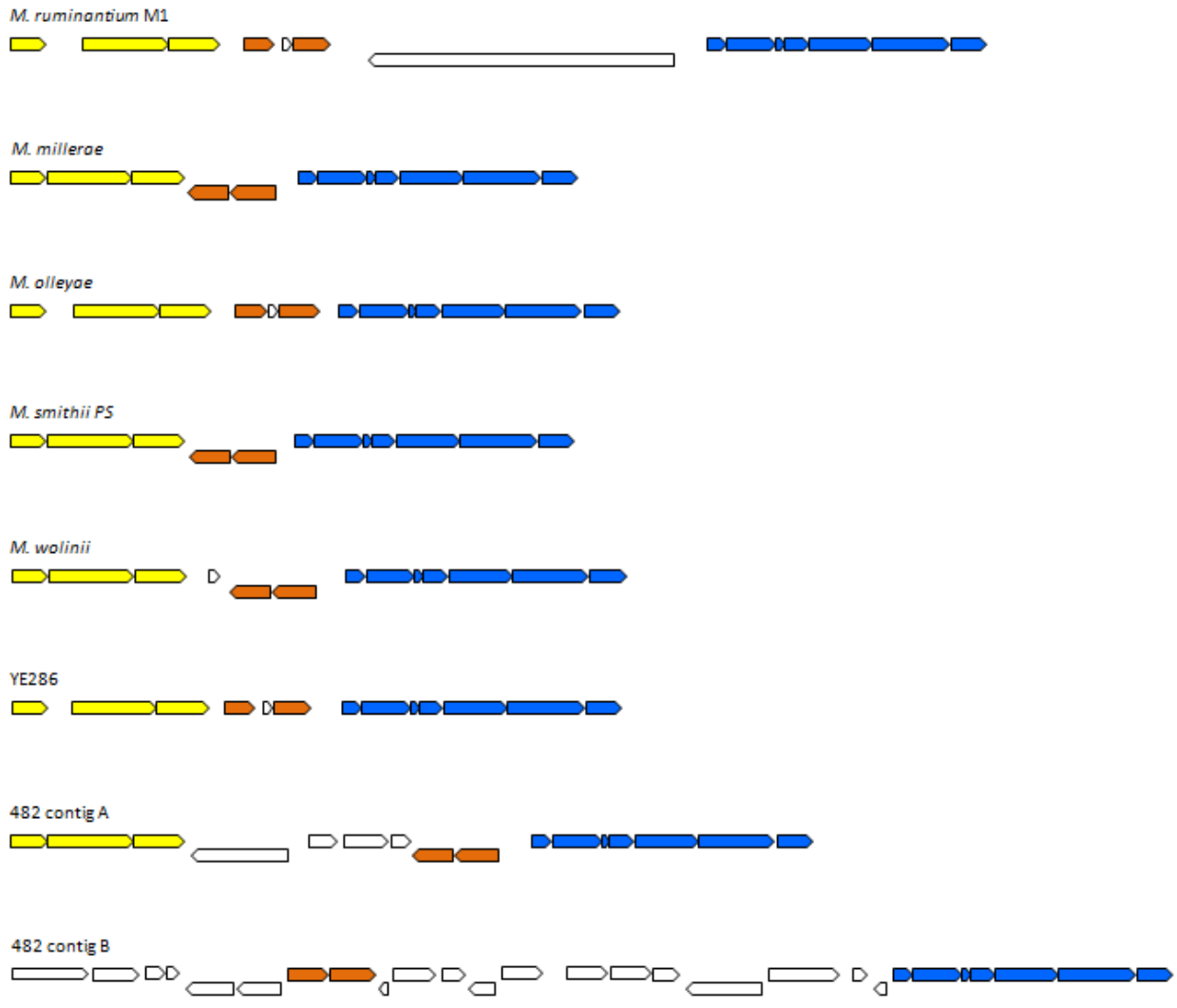


Figure 10. Genomic architecture of formate dehydrogenase operons (fdhCAB, yellow) and formylmethanofuran dehydrogenase operons (fwdCABDGFH, blue) in the genomes of selected *Methanobrevibacter* strains the rumen methanogen metagenome database. Orange genes encode molybdopterin-guanine dinucleotide biosynthesis proteins that appear to be collocated in this genomic region. 482 contig A and 482 contig B have been selected as examples for this figure - the information presented here is not intended to represent all of the fdh and fwd operons found in the metagenome database

The ability of some methanogens to utilise multiple substrates may provide the metabolic flexibility required to survive under different conditions. The *M. wolinii* SH genome encodes loci for formate and methanol utilisation, yet these substrates are not used for methanogenesis when the microbe is grown in a monoculture (Miller and Lin, 2002). Although it is not clear whether these pathways are functional in this strain, evidence from other *Methanobrevibacter* isolates suggests that substrate utilisation may be related to the activities of other microbes within the gut. When *M. smithii* PS is introduced into gnotobiotic (germ-free) mice following colonisation with the formate and methanol producing bacterium *Bacteroides thetaiotamicron*, expression of the formate and methanol utilisation genes are upregulated (Samuel *et al*., 2007). Similarly, genes in the formate dehydrogenase operon are upregulated when *M. ruminantium* M1 is grown in co-culture with the formate producing bacterium *Butyrivibrio proteoclasticus* B316 (Leahy *et*

al., 2010). This may suggest that formate and/or methanol are only used when interspecies substrate transfer can occur. These so-called syntrophic interactions are known to be important for the persistence of many fermentative bacteria in the gut environment, preventing build-up of metabolic end products that can inhibit further growth. It has been shown in gnotobiotic mice that the presence of *M. smithii* can shift *B. thetaiotamicron* polysaccharide degradation away from production of propionate and butyrate towards production of formate and acetate, with the latter used by *M. smithii* as a source of cell carbon (Samuel and Gordon, 2006). At the microbiological level, it is important to understand how different substrates are produced and used under *in vivo* (rumen) conditions when considering targeted strategies for reducing methane emissions from livestock.

Characterisation of rumen methanogen gene expression in northern Australian beef cattle

The database of rumen methanogen functional genetic potential provides a catalogue of "what they can do". It is essentially a description of mechanisms that can be used for survival and proliferation under different *in vivo* conditions. Metatranscriptomics uses next-generation sequencing to determine which genes in a sample are being expressed. The ability to determine gene expression profiles of methanogens in rumen samples allows us to answer the more important question of "what are they doing?". Up until now this research question has been difficult to answer for methanogens in rumen samples, as it relies on the use of genome sequence data to act as a "scaffold" upon which a picture of gene expression can be built. For the first time in this project it has been possible to produce genome sequence data from the predominant *Methanobrevibacter* and *Methanosphaera* species in ruminants using a novel culture-independent approach. We generated a second database of rumen methanogen functional potential from steers consuming a diet of improved Rhodes grass pasture +/- supplementation with *Leucaena* (in collaboration with 01200.035/B.CCH.6510). The aim of this study was to determine if it is possible to use the metagenome database as a scaffold to study methanogen gene expression *in vivo*, and to describe the mode of action of *Leucaena* as a methane mitigation strategy at the microbiological level. A diagrammatic overview of the method is provided in Figure 11.

The protocol for increasing representation of methanogen DNA was applied to rumen fluid samples from ten steers consuming Rhodes grass +/- *Leucaena* (n=5 per group). Samples were subjected to next-generation sequencing, which returned an average of 7.36×10^7 reads per sample (101 base pairs each). The ribosomal RNA reads were extracted using RNammer and 16S rRNA reads were classified using the Ribosomal Database Project (RDP) hierarchy. Classification indicated that 20-25% of the 16S rRNA reads from the *Leucaena* supplemented steers were archaeal in origin, and that this increased to 30-39% for the steers consuming only Rhodes grass (Figure 12). The metagenomes from the *Leucaena* supplemented steers showed an increased representation of data from the *Methanosphaera* genus, comprising 11-16% of the total archaeal reads (compared to 5-6% for the Rhodes grass only fed steers; Figure 12).

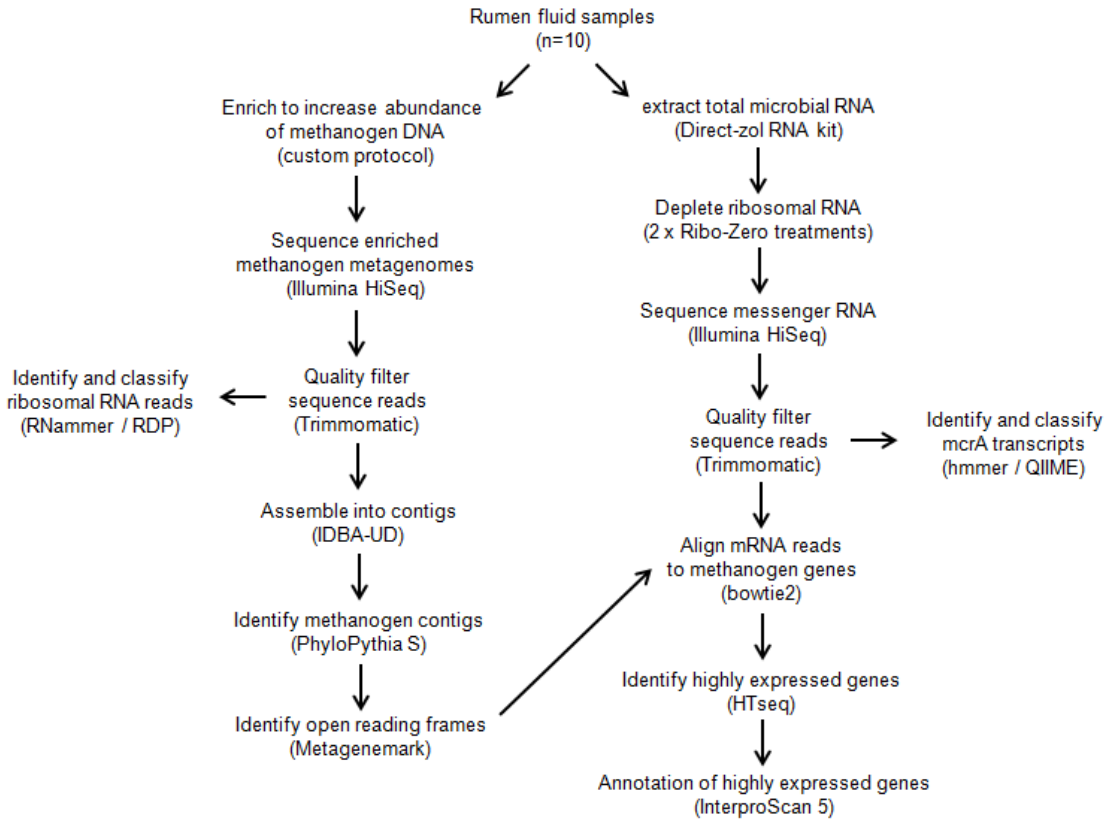


Figure 11. Flowchart depicting the methods used to analyse methanogen gene expression in beef cattle consuming Rhodes grass +/- supplementation with *Leucaena*

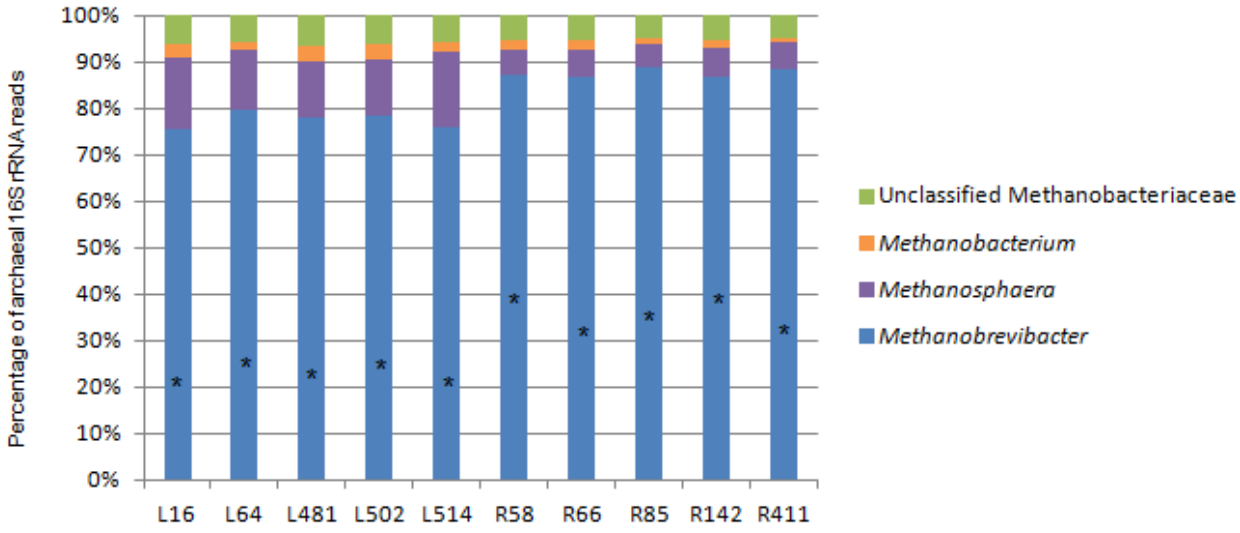


Figure 12. Classification of archaeal 16S rRNA reads from rumen samples treated to increase the representation of methanogen DNA. Samples prefixed with "L" represent steers fed Rhodes grass + *Leucaena*; samples prefixed with "R" represent steers fed Rhodes grass only. Asterisks indicate the percentage of archaeal 16S rRNA reads relative to bacterial 16S rRNA reads for each sample

The data was processed in the same way as previously described for the Flinders grass metagenomes. Reads from individual samples were assembled separately using IDBA-UD (Table 4; Peng *et al*., 2012), before using the PhyloPythiaS online web server to classify large contigs from each assembly into domain level taxonomic "bins", thereby separating the archaeal contigs from the bacterial contigs (Table 5) and creating a methanogen metagenome from each animal. This process provided 10 datasets, each containing 36-81 Mb of data. In total, the rumen methanogen metagenome from this cohort of cattle contains almost 600,000 genes from the *Methanobrevibacter* and *Methanosphaera* genera. Since the metagenomes have been assembled separately for sample, many of these genes will have been replicated as the same methanogens are likely to be present in the rumens of all cattle in the cohort.

Table 4. Assembly data for each metagenome in the Rhodes grass +/- *Leucaena* study (before separation of bacterial and archaeal contigs)

Sample (steer number)	Number of trimmed input reads	Number of large contigs (> 1Kb)	Size of assembled dataset (large contigs; Mb)	% of input reads used in assembly (large contigs)
L16	7.15×10^7	84814	205.0	48.2
L64	8.01×10^7	92288	233.0	51.0
L481	8.03×10^7	92038	216.0	46.9
L502	8.03×10^7	91631	238.9	49.1
L514	5.93×10^7	61402	174.7	51.1
R58	7.07×10^7	78739	210.6	56.8
R66	7.76×10^7	88031	229.1	53.2
R85	7.71×10^7	81247	215.6	58.1
R142	6.71×10^7	77501	196.9	54.8
R411	7.17×10^7	74975	186.1	52.0

Table 5. Properties of enriched methanogen metagenome datasets from the Rhodes grass +/- *Leucaena* study

Sample (steer number)	Number of archaeal contigs	Mean length/N50 (bp)	Total size (Mb)	% of input reads aligned to contigs	Total number of genes
L16	6795	5321 / 9227	36.2	13.9	35870
L64	12766	5092 / 7900	65.0	23.1	64266
L481	9436	4944 / 8412	46.7	15.5	46190
L502	12178	5353 / 8461	65.2	20.4	63281
L514	6753	6143 / 10442	41.5	17.4	40495
R58	14943	5414 / 8199	80.9	34.8	77486
R66	12846	5631 / 8674	72.3	26.9	70697
R85	12899	5799 / 9112	74.8	32.6	71788
R142	14316	4933 / 7484	70.6	32.6	68739
R411	11202	5117 / 8020	57.3	28.0	55722

The next step in the analysis was to use the enriched metagenomes as a template for detecting methanogen gene expression in these samples. Matched metatranscriptomes were prepared and sequenced on two lanes of Illumina HiSeq 2000, providing an average of 6.1×10^7 trimmed reads per sample. The percentage of ribosomal RNA reads (as determined using RNammer) was between 0.8 and 2.5%, indicating that treatment of total RNA with two rounds of Ribo-Zero (as described in the methods section) is an efficient way to increase the amount of messenger RNA ("informative signal") in a rumen microbiome sample prior to sequencing.

Messenger RNA reads were aligned to the matching methanogen metagenome contigs for each sample using bowtie 2 (Langmead and Salzberg, 2012), with highly expressed genes identified using HTseq (Anders *et al*., 2014). The top 50 highly expressed genes for each sample were annotated with Blast2GO (Gotz *et al*., 2008) and InterproScan 5 (Jones *et al*., 2014). No significant differences were observed between the two diets, with expressed genes assigned to members of the genus *Methanobrevibacter*. The analysis highlighted enzymes and transporters associated with the central methanogenesis pathway, including formate dehydrogenase, suggesting that formate is an important substrate for methane production in this cohort of animals. We identified systems for cobalt uptake and cobalamin biosynthesis, essential for the function of the membrane bound enzyme Mtr which catalyses the penultimate step in reduction of carbon dioxide with hydrogen to form methane. Several genes predicted to encode adhesin-like proteins were also revealed. Adhesins are large cell-surface exposed proteins that enable attachment to surfaces or other cells. Based on the results from our complete annotation of the pilot rumen methanogen metagenome database (from the 3 Brahman steers fed Flinders grass), we estimate that adhesin-like proteins represent approximately 6% of genes in an average *Methanobrevibacter* genome. With each metagenome containing approximately 6 x 10^4 genes (Table 5), we sequenced around 3,600 adhesin-like proteins per sample. The metatranscriptomic analysis has allowed us to identify and prioritise those that are used by the most active methanogens in rumen samples.

Results from project 01200.035/B.CCH.6510 indicate that a substantial proportion of the rumen methanogen populations in this cohort of beef cattle are represented by the Methanomassiliicoccales. The methanogen population structure of the 10 samples included in the current analysis is provided in Figure 13, showing that the relative percentage of Methanomassiliicoccales ranges from 19 to 32%. This is in contrast to the results presented earlier in this report, where the Methanomassilicoccales did not comprise a large proportion of the methanogen populations in beef cattle consuming a range of tropical forages (Figure 5; range 2 to 11%) or in sheep consuming oaten chaff +/- supplementation with *Eremophila* (Figure 6; range 1 to 4%).

The amplicon profiling data indicates that members of the *Methanobrevibacter* genus are the predominant methanogens in this cohort of beef cattle from the Rhodes grass +/- *Leucaena* study, but that members of the Methanomassiliicoccales may also be significant contributors to the total yield of methane produced. Our method for generating the rumen methanogen metagenome database did not capture any genome sequence information from the Methanomassiliicoccales, as seen by contrasting the population structures shown in Figures 12 and 14. In order to get a more complete picture of the metabolic activity of all methanogens in these rumen samples, we identified mRNA sequence reads generated by transcription of the methyl coenzyme M reductase subunit A (mcrA) gene. This was achieved using hmmer (Eddy, 2011) and two hidden Markov models from the Pfam database (MCR_alpha_N: PF02745; and MCR_alpha: PF02249; Finn *et al*., 2014) that represent conserved motifs found in all mcrA sequences (Figure 14). This method is similar to the use of RNammer (Lagensen *et al*., 2007) for identification of 16S rRNA gene fragments in metagenomic sequencing runs. Since we

deliberately depleted transcripts derived from ribosomal RNA genes prior to sequencing (otherwise the signals from messenger RNA would have been difficult to detect), we have instead used the mcrA gene as a "marker" to determine which methanogens are metabolically active in each sample. The mcrA gene is useful as an informative phylogenetic marker, as methyl coenzyme M reductase catalyses the final step in the production of methane for all known methanogens.

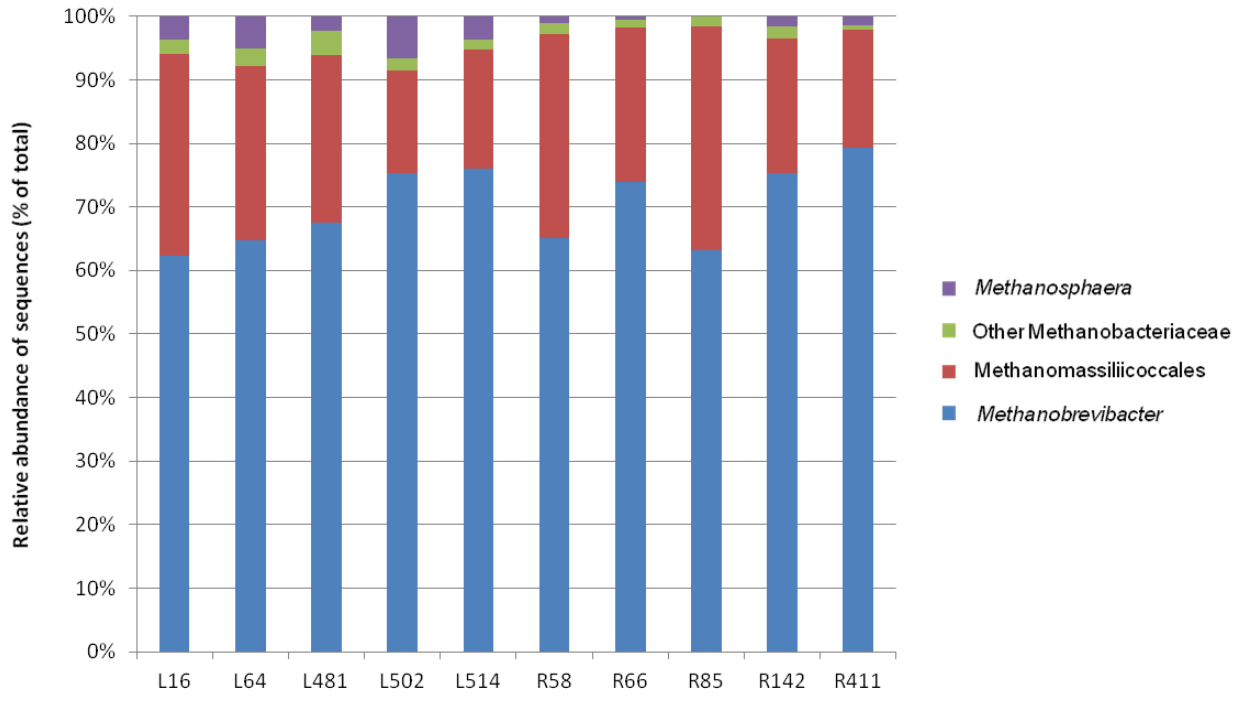


Figure 13. Composition of rumen methanogen populations in beef cattle consuming Rhodes grass +/- supplementation with *Leucaena*. Results were generated as part of project 01200.035/B.CCH.6510 from archaeal-specific 16S rRNA gene amplicon profiling using a protocol similar to that described in the methods section of this report

Figure 14. Methyl-coenzyme M reductase subunit A (mcrA) gene from *Methanobrevibacter ruminantium* M1 (1656 base pairs), showing the nucleotide position of hidden Markov models MCR_alpha_N (PF02745; bases 10-804) and MCR_alpha (PF02249; bases 943-1323)

On average, we retrieved 2 x 10^4 mcrA transcripts from each metatranscriptomic dataset using the two hidden Markov models, which cover approximately 70% of the gene. The transcripts were compared to a library of full-length mcrA genes using closed reference OTU picking in QIIME (Caporaso et al., 2010) at 92% similarity cutoff. Approximately 95% of the mcrA transcripts from each sample were successfully matched to a reference gene in the library. The library of mcrA genes (n=111) included sequences from the two methanogen metagenome databases (n=50); sequences from mcr operons assembled as part of

a study to measure expression of selected methanogen genes in the rumen of low and high methane emitting sheep (n=33; Shi *et al*., 2014); and sequences from selected cultured representatives of several genera including *Methanobrevibacter*, *Methanosphaera*, Methanomassiliicoccales, *Methanomicrobium* and *Methanosarcina* from gut microbiomes (n=28; available at http://img.jgi.doe.gov or as described by Shi *et al*., 2014).

There are two isoforms of the methyl-coenzyme M reductase enzyme, denoted as I and II (or mcr and mrt, respectively). The genomes of many methanogens (including some members of the *Methanobrevibacter* genus) encode both enzyme isoforms, thought to enable a rapid response to changes in hydrogen concentration (Reeve *et al*., 1997). Members of the *Methanosphaera* and Methanomassiliiococcales appear to encode the mrt isoform, as described by Shi *et al* (2014). For eight of the nine samples analysed, the majority of transcripts were most similar to reference genes belonging to the *Methanobrevibacter* mcrA cluster, irrespective of dietary supplementation with *Leucaena* (Figure 16). This indicates that methanogen gene expression as determined by this experimental protocol is dominated by transcripts from members of the *Methanobrevibacter* genus. One sample (L64) exhibits a different transcript profile, where almost 60% of reads were assigned to an mrtA gene. The cause of this result is unclear.

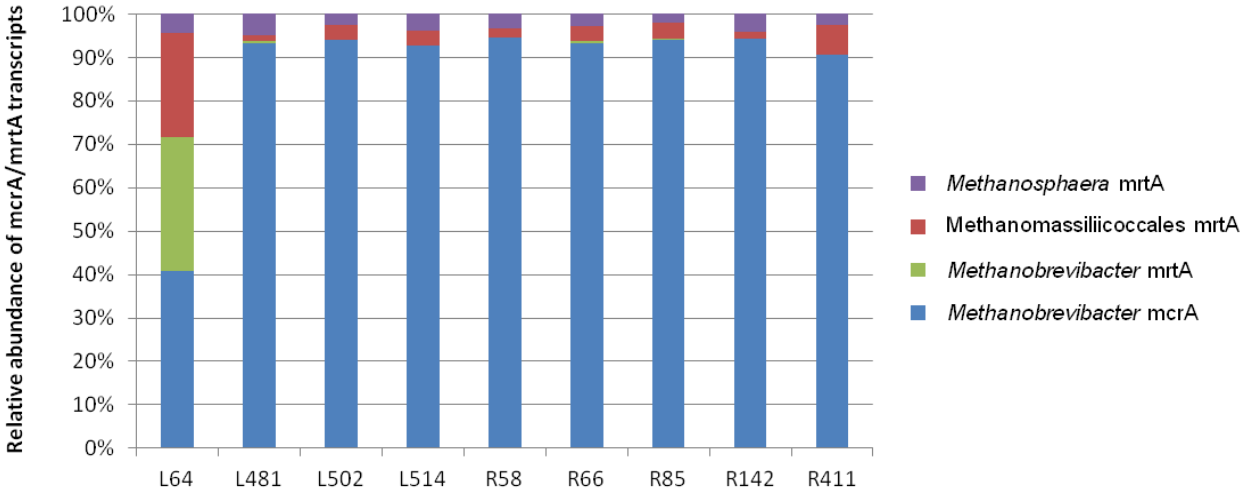


Figure 15. Assignment of methyl-coenzyme M reductase A transcripts from beef cattle rumen samples to a functional phylogenetic grouping, based an expanded database of mcrA/mrtA genes obtained from this study, from selected genomes available at http://img.jgi.doe.gov and the detailed clade analysis presented by Shi *et al* (2014)

4. Discussion

Microbial genomics involves the study of genes and their associated functions by combining genetic sequencing with computational bioinformatics. It provides a means to analyse the various cellular mechanisms, nutrient requirements and niche adaptations used by microbes to survive and persist under different environmental conditions. At the commencement of the current project, only one genome

sequence from a rumen methanogen had been made available in a public data repository (Leahy *et al*., 2010). The lack of genome sequence data has proved to be rate-limiting for development of targeted strategies to inhibit methanogens in ruminants. The overarching objective of this project was to define the functional metabolic potential of rumen methanogens and provide new information about the systems critical for their proliferation in Australian production systems. This was achieved by implementing a suite of novel tools to:

(i) highlight the predominant methanogen strains in selected cohorts of cattle and sheep ("who's there?")
(ii) catalogue the functional metabolic potential of the predominant strains in a rumen methanogen metagenome database ("what are they capable of?")
(iii) use the metagenome database as a template to determine which genes are being expressed under *in vivo* conditions ("what are they doing?")
(iv) identify conserved cell-surface exposed proteins as targets for methane mitigation strategies ("how can we stop them?")

Through this research project we have filled a significant gap in our understanding of "what makes a rumen methanogen tick" and have presented some practical suggestions for how this information can be used to reduce methane emissions from livestock. These are summarised in the following sections.

Who's there?

Our analyses of rumen methanogen community structure in sheep and cattle indicated that members of the *Methanobrevibacter* genus are the predominant methanogens found in Australian production systems. As a result, we chose to focus all subsequent research on characterising the functional metabolic potential of this genus in rumen fluid samples. Traditional approaches that rely on laboratory cultivation are time-consuming and don't always result in isolation of strains relevant to methane production *in vivo*. For example, the best characterised rumen methanogen is *Methanobrevibacter ruminantium* M1 - we showed that this strain is only representative of a small proportion of the methanogens present in the rumen samples we surveyed. A different approach was required in order to catalogue the genetic potential of the *Methanobrevibacter* genus at a population scale.

What are they capable of?

In a typical rumen fluid sample, methanogens only comprise a small proportion of the total prokaryotic cells (approximately 1%). Standard approaches to study the DNA content of abundant microbes in a sample ("shotgun metagenomics") would not be able to provide sufficient data to facilitate a critical evaluation of rumen methanogen genomic content - it is akin to searching for a needle in a haystack. We successfully implemented a novel protocol to increase the relative abundance of genetic material from methanogens in DNA extracted from rumen samples. We generated a pilot rumen methanogen metagenome database from three Brahman steers fed Flinders grass. It is a comprehensive resource containing 1.2×10^5 annotated genes from members of the *Methanobrevibacter* and *Methanosphaera*

genera. It represents the first description of the genetic repertoire of these methanogens at a population level, using a method that reflects the predominant strains relevant to methane production by Australian livestock.

The rumen supports a complex microbial community that breaks down plant biomass, yielding proteins and short-chain fatty acids that can be used by the animal. Methanogens use some of the simple carbon compounds produced by microbial fermentation as terminal electron acceptors, and as such they represent the "end of the line" for metabolism in the rumen - methane is not used by other microbes in this system and is lost to the atmosphere by eructation. Interrogation of the pilot database has revealed that rumen *Methanobrevibacter* strains have the ability to utilise formate in addition to carbon dioxide as a significant source of terminal electron acceptors. An understanding of the substrate utilisation capacity of rumen methanogens will provide opportunities to redirect fermentation towards schemes that are less conducive to their growth, reducing methane emissions and increasing the potential for productivity gains in the animal through improved carbon capture.

What are they doing?

Our strategy to catalogue the functional metabolic potential harboured by rumen members of the order Methanobacteriales has provided us with a wealth of information to describe the systems that these methanogens are capable of using for survival and growth. The metagenome database is analogous to a playbook describing all possible strategies that could be used during a sports game. The actual combination of mechanisms employed by methanogens at any given point in time will be contingent on conditions in the rumen, just the same as a sports team would select different plays depending on the circumstances. As a fan of the game it would be much easier to interpret the plays being employed if you had a copy of the playbook describing the signals and strategies, rather than relying on simply observing the outcome. In a similar way we have been able to describe exactly the activities of the predominant methanogens under *in vivo* (rumen) conditions by comparing gene expression data ("the plays") to the metagenome database ("the playbook"). It provides a tangible link between "who's there" and "what they are doing", revealing the actual mechanisms being used by the methanogens to successfully colonise the rumen microbiome.

We prepared paired metagenomic and metatranscriptomic datasets from beef cattle fed Rhodes grass +/- supplementation with *Leucaena*, in collaboration with project 01200.029/B.CCH.6510. Strains from the *Methanobrevibacter* genus represented the predominant methanogens in both datasets, providing a unique resource to describe their functional genetic potential and patterns of gene expression in a relevant Australian livestock production system. To our knowledge, this is the first example of a study that has successfully detected highly expressed genes of rumen methanogens *in vivo* using population scale metagenomic data as a scaffold to guide the analysis. By taking this approach the datasets themselves have revealed which genes are being transcribed. As expected, the list of highly expressed genes includes proteins directly involved in reduction of carbon dioxide with hydrogen to methane and the

associated energy conservation reactions. The remaining genes in the list can therefore be considered to encode proteins essential for methanogen growth at the time of sampling as their expression levels were also determined to be high. Three important observations pertaining to the activity of methanogens in this system are described in the following paragraphs.

(i) using formate as a substrate for methanogenesis

In a system with high turnover such as the rumen, some products of fermentation will be used as quickly as they are produced. This has been demonstrated for succinate, which is rapidly converted to propionate by bacteria (Blackburn and Hungate, 1963) and hydrogen, used for methanogenesis (Hungate, 1967). Evidence indicates that methanogenic substrates produced by fermentative microbes (bacteria, protozoa and fungi) can be directly transferred to methanogens via interspecies substrate transfer. These factors make it difficult to assess how much formate is being used for methanogenesis with biochemical measures (Hungate *et al*., 1970). Formate dehydrogenase genes were highly expressed in rumen samples from beef cattle fed Rhodes grass +/- supplementation with *Leucaena*, suggesting that formate is an important precursor for methane formation on both diets. Monitoring expression of formate dehydrogenase genes is a simple way to establish the relevance of formate for methane production in other systems

(ii) transporting cobalt and synthesising cobalamin

Cobalt ions are used to synthesise cobalamin (coenzyme B_{12}), a cofactor for the membrane associated Mtr complex which catalyses an essential step in the central methanogenesis pathway for members of the *Methanobrevibacter* genus (conversion of methyl H_4MPT to methyl CoM; see Figure 1). The list of highly expressed genes of rumen *Methanobrevibacter* strains *in vivo* includes subunits of the Mtr complex, ABC transport systems for the uptake of cobalt and components of the cobalamin biosynthetic pathway. The majority of proteins encoded by the upregulated genes for cobalt uptake are conserved in methanogens and many lineages of rumen bacteria, and as such they do not represent viable targets for development of vaccine-based methane mitigation strategies. The results do support our initial hypothesis surrounding the importance of transition metals in the life cycle of rumen methanogens. Dietary supplementation with cobalt to improve productivity or rectify an insufficiency may therefore have the unintended consequence of increasing ruminant methane emissions under certain circumstances.

(iii) expressing a subset of adhesin-like proteins

Adhesins are cell-surface exposed proteins that allow bacteria to attach to other microbes or surfaces. Adhesin-like proteins (ALPs) are common in the genomes of gut Methanobacteriales strains, and it has been suggested that regulated expression of ALPs may promote syntrophic relationships with substrate-producing microbes (Samuel *et al*., 2007; Leahy et al., 2010; Hansen *et al*., 2011). Based on results from annotating the pilot methanogen metagenome database, we estimate that 6% of genes encode putative ALPs. Given that this represents a significant proportion of the dataset, we needed a means to prioritise those that are functionally relevant for further analysis. Several ALPs were highly expressed in the Rhodes grass +/- *Leucaena* rumen samples. Many contain features indicating involvement in the

physical attachments between methanogens and other rumen microbes that enable interspecies substrate transfer to occur. Disruption of these interactions may provide a novel means to negatively impact the growth potential and/or rate of methanogenesis in ruminant livestock.

The metatranscriptome data has highlighted a selection of genes that are highly expressed in a subset of rumen samples (n=9) at the time of collection. This has provided us with a set of targets that still need to be validated in a larger sample set (using quantitative PCR) in order to confirm our preliminary findings. We are expecting to observe some variation in the results (as was observed for the metatranscriptome for steer L64, refer to Figure 15) that are not necessarily related to supplementation with *Leucaena*. The cohort of mixed-breed cattle was maintained under natural grazing conditions where they were able to roam freely and consume fresh forage *ad libitum*. Diurnal fluctuations in ruminant methane production are related to feeding patterns, as the availability of methanogenic substrates is dependent on microbial fermentation of ingested plant biomass. As a result, methanogen gene expression profiles will particularly impacted by the time since their last feeding. We deliberately chose to collaborate with project 01200.029/B.CCH.6510, as we saw advantages in studying a model production system for northern Australian beef where methane emissions are typically high and options for mitigation are limited. Given the potential limitations, it is encouraging to see patterns emerging from the initial analyses that provide an insight into rumen methanogen growth and methane formation under “real world” conditions.

How can we stop them?

The metagenome database provides a detailed description of the functional genetic potential inherent to members of the *Methanobrevibacter* and *Methanosphaera* genera in the rumen microbiome. We have translated these results into a precise focus (rather than an empirical “shotgun” approach) for efforts to produce methanogen inhibitors. A comprehensive analysis of the database revealed a suite of cell-surface exposed proteins from predominant methanogens found in Australian production systems. We have documented the sequence variation in selected “key” proteins at a level of detail and depth that has not been achieved previously. This will form the basis for ongoing studies seeking to design vaccines for use in ruminants, providing essential information to help ensure that the methanogens responsible for methane production *in vivo* are being targeted by the strategies under development.

Concluding remarks

In this project, we set out to fill a basic research gap surrounding the mechanisms used by methanogens to survive and proliferate in the rumen. The paucity of knowledge to describe their functional metabolic activity was rate-limiting for efforts to control methane production from livestock using targeted approaches. We have been able to reveal “critical control points” of rumen methanogenesis that can now be exploited in future research. This will ultimately benefit Australian agriculture by providing new pathways to inhibit methane production in ruminants based on a detailed understanding of the microbes involved. The novel techniques we have developed are also applicable to studies seeking to understand the dynamics of methanogenesis for other agricultural applications (e.g. manure management). The tools

described in this report will be useful for quantifying the impact of any abatement strategy on rumen methanogen populations. We have shown how it can provide useful information about the system (e.g. in terms of methanogen substrate utilisation preferences) that would have been difficult to obtain by traditional measures. This insight can then be used to redirect the rumen microbiology towards fermentation schemes that are less conducive to production of methanogenic substrates, with the aim of reducing methane emissions and increasing productivity through improved carbon capture within the animal.

5. Future research needs

Expansion of the rumen metagenome database

Our ability to successfully develop targeted approaches for mitigating ruminant methane emissions is highly dependent on building a fundamental understanding of the underlying microbiology. The research described in this report has made significant advances in this space by characterising the functional metabolic potential of rumen Methanobacteriales by sequencing genomes of phenotypically relevant methanogens and identifying key genes essential for survival and growth. The methods we devised were specifically designed to enable study of the *Methanobrevibacter* genus, but were also able to provide genomic data from the *Methanosphaera* genus. Not all methanogens are created equal, however, and a different approach is required in order to obtain a more complete picture.

Several studies have described the recently proposed 7th order of methanogens known as the Methanomassiliicoccales (Paul *et al*., 2012; Iino *et al*., 2013; Poulsen *et al*., 2013; Borrel *et al*., 2014; Lang *et al*., 2015). They exhibit unique metabolic properties compared to other strict hydrogen-dependent methanogens, characterised by the ability to utilise methanol, methylated amines or dimethyl sufides as substrates for growth and the absence of a coenzyme $F4_{20}$-depedent methanogenesis cascade (Borrel *et al*., 2014). Unlike the Methanobacteriales, which possess thick cell walls resistant to permeabilisation and lysis (Kubota *et al*., 2007), the Methanomassiliicoccales appear to lack a proper cell wall (Lang *et al*., 2015) and are sensitive to freeze-thawing and osmotic stress. Rumen samples need to be treated carefully to ensure all cells remain intact prior to nucleic acid extraction or the contribution of the Methanomassiliicoccales to methane production may be underestimated. Likewise, extraction methods that do not enable efficient cell lysis may underestimate the contribution of the Methanobacteriales. This should be taken into account by anyone seeking to analyse methanogens in rumen samples.

There is a growing body of evidence that suggests that a focused study to define the genetic potential of the Methanomassiliicoccales in ruminants is warranted. A reduction in methane output from dairy cows fed rapeseed oil was correlated with a decrease in expression of key genes belonging to the Methanomassiliicoccales (Poulsen *et al*., 2013). Transcriptional upregulation of key genes belonging to both the *Methanobrevibacter* sp. and the Methanomassiliicoccales was shown to underpin increases in

methane emissions from sheep selected for a high emission phenotype (Shi *et al*., 2014). Despite the recognition that the Methanomassiliicoccales are an important component of the rumen microbiota, representative genome sequence data is absent from public repositories. Based on the limited information that is available (from published human gut isolates as summarised by Borrel *et al* (2014) and in-house data from selected macropodid, chicken and rumen isolates), the genomes that we have studied contain many genes that are specific to methanogens from this order. Similarly, many of the "key" proteins identified in our analysis of the Methanobacteriales metagenome database are not found in sequenced representatives of gut Methanomassiliicoccales available in the IMG database (n=4). Taken together, this suggests that ongoing genomics-based efforts to develop targeted antimethanogenic vaccines or chemical inhibitors would benefit from ruminant-specific Methanomassiliicoccales (meta)genome sequence data to complement the resources already produced over the course of the current project.

Symbioses between rumen protozoa and prokaryotes

Methanogens can be found as free-living cells in the rumen fluid, or they can be physically attached to protozoa or bacteria where they participate in interspecies hydrogen transfer. The latter is a mutually beneficial symbiosis where excess hydrogen from fermentation is removed and used for methanogenesis. Many studies have identified methanogens associated with different species of rumen protozoa using community profiling techniques (for recent examples, see Tymensen *et al*., 2012 and Belanche *et al*., 2014). Such studies are underpinned by relatively simple methods to obtain protozoal preparations that are free from contamination with free-living prokaryotes (bacteria and methanogens). It has been estimated that 37% of methane in ruminants is generated by protozoa-associated methanogens (Finlay *et al*., 1994). A targeted study to determine how methanogens are able to facilitate attachment to protozoa (e.g. by expression of adhesin-like proteins) may present novel pathways for methane mitigation.

Results from project 12000.029/B.CCH.6420 have shown that the red macroalga *Asparagopsis taxiformis* is a direct and potent inhibitor of methanogenesis in rumen fluid. We collaborated directly with Lorenna Machado (PhD student, James Cook University) and Nigel Tomkins (CSIRO) to confirm that bromoform ($CHBr_3$) is the active component of *A. taxiformis*. Addition of 2% *A. taxiformis* or 5 μM bromoform to an *in vitro* fermentation of Rhodes grass results in a 98% decrease in methane output with a concomitant reduction in methanogen relative abundance and analogous changes in bacterial community composition that are related to increases in hydrogen concentration (Machado *et al*., manuscript in preparation). Since bromoform appears to be inhibitory to all methanogens, it is reasonable to suggest that it will be active against the protozoa-associated fraction. Provided that dietary supplementation with *A. taxiformis* does not adversely impact the protozoal population (no significant differences were detected in the *in vitro* fermentation), it may provide an opportunity to study alternate pathways for hydrogen utilisation. Specialised bacteria such as homoacetogens, which use carbon dioxide and hydrogen to make acetate, would normally be unable to occupy this niche as they would be outcompeted by methanogens.

An analysis of rumen samples from animals fed a standard diet +/- supplementation with *A. taxiformis* using microbial community profiling would provide confirmation of the hypothesised mode of action *in vivo*. A concurrent assessment of protozoa and their associated prokaryotic symbionts would enable the additional objectives described in the preceding paragraphs to also be achieved. This research need could be achieved by a PhD project carried out in conjunction with the upcoming planned feedlot trials for *A. taxiformis*, or as a smaller separate study.

6. Publications and websites

- Rosewarne CP, Bradbury MI, Denman SE and McSweeney CS (2015) Searching for methanogen genomic needles in a rumen microbiome haystack. Poster presented at the Genomic Standards Consortium Meeting (GSC 17) in Walnut Creek, CA; and the American Society for Microbiology General Meeting (ASM 2015) in New Orleans, LA. Refer to Appendix 3

- CSIRO Agriculture Flagship http://www.csiro.au/en/Research/AF

- Carly Rosewarne, Project Leader http://people.csiro.au/R/C/Carly-Rosewarne.aspx

7. Reference list

- Anders and Huber (2010) Differential expression analysis for sequence count data. Genome Biology 11:R106
- Anders *et al* (2015) HTSeq - a python framework to work with high-throughput sequencing data. Bioinformatics 31(2): 166-169
- Aguirre *et al* (2011a) Strategy for modular tagged high-throughput amplicon sequencing. Applied and Environmental Microbiology 77(17): 6310-6312
- Aguirre *et al* (2011b) Evaluation of subsampling-based normalization strategies for tagged high-throughput sequencing data sets from gut microbiomes. Applied and Environmental Microbiology 77(24): 8795-8798
- Attwood *et al* (2011) Exploring rumen methanogen genomes to identify targets for methane mitigation strategies. Animal Feed Science and Technology 166-167: 65-75
- Bailey *et al* (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research 37: W202-208
- Belanche *et al* (2014) Study of methanogen communities associated with different rumen protozoal populations. FEMS Microbiology Ecology 90: 663-677
- Blackburn and Hungate (1963) Succinic acid turnover and propionate production in the bovine rumen. Applied Microbiology 11:132-135
- Bolger *et al* (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15): 2114-2120
- Caporaso *et al* (2010) QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7: 335-336
- Cole *et al* (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Research 37: D141-145
- Denman *et al* (2007) Quantitation and diversity analysis of ruminal methanogenic populations in response to the antimethanogenic compound bromochloromethane. FEMS Microbiology Ecology 62: 313-322
- Denman *et al* (2006) Development of a real-time PCR assay for monitoring anaerobic fungal and cellulolytic bacterial populations within the rumen. FEMS Microbiology Ecology 58: 572-582
- Dereeper *et al* (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Research 36: W465-459
- Dobson *et al* (1956) A histological study of the organisation of the rumen epithelium in sheep. Quarterly Journal of Experimental Physiology and Cognate Medical Sciences 41: 247-253
- Eddy (2011) Accelerated profile HMM searches. PLoS Computational Biology 7:e1002195
- Edgar (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature Methods 10: 996-998
- Finlay *et al* (1994) Some rumen ciliates have endosymbiotic methanogens. FEMS Microbiology Letters 117: 157-162
- Finn et al (2014) The Pfam protein families database. Nucleic Acids Research 42:D222-D230
- Gantner *et al* (2011) Novel primers for 16S rRNA-based archaeal community analyses in environmental samples. Journal of Microbiological Methods 84(1): 12-18
- Fricke *et al* (2006) The genome sequence of Methanosphaera stadtmanae reveals why this human intestinal archaeon is restricted to methanol and H_2 for methane formation and ATP synthesis. Journal of Bacteriology 188(2): 642-658
- Gotz *et al* (2008) High throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research 36: 3420-3435
- Gottschalk and Thauer (2001) The Na+-translocating methyltransferase complex from methanogenic archaea. Biochimica et Biophysica Acta - Bioenergetics 1505(1): 28-36

- Grüber *et al* (2014) ATP synthases from archaea: the beauty of a molecular motor. Biochimica et Biophysica Acta - Bioenergetics 1837(6): 940-952
- Hansen *et al* (2011) Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. Proceedings of the National Academy of Sciences 108(1): 4599-4606
- Hungate (1967) Hydrogen as an intermediate in rumen fermentation. Archiv für Mikrobiologie 59:158-164
- Hungate *et al* (1970) Formate as an intermediate in the bovine rumen fermentation. Journal of Bacteriology 102(2): 389-397
- Janssen and Kirs (2008) Structure of the archaeal community of the rumen. Applied and Environmental Microbiology 74(12): 3619-3625
- Jones *et al* (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30 (9): 1236-1240
- Kaster *et al* (2011) Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. Proceedings of the National Academy of Sciences 108(7): 2981-2986
- Kennedy and Charmley (2012) Methane yields from Brahman cattle fed tropical grasses and legumes. Animal Production Science 52(4): 225-239
- Kittlemann *et al* (2013) Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. PLoS One 8(2): e47879
- Kubota *et al* (2007) Evaluation of enzymatic treatments for application of CARD-FISH to methanogens. Journal of Microbiological Methods 72(1): 54-59
- Krogh *et al* (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of Molecular Biology 305(3): 567-580
- Iino *et al* (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 and Environments 28(2): 244-50
- Lagensen *et al* (2007) RNammer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Research 35(9): 3100-3108
- Lang *et al* (2015) New mode of energy metabolism in the seventh order of methanogens as revealed by comparative genome analysis of “Candidatus Methanoplasma termitum”. Applied and Environmental Microbiology 81(4): 1338-1352
- Langmead and Salzberg (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357-359
- Leahy *et al* (2010) The genome sequence of the rumen methanogen *Methanobrevibacter ruminantium* M1 reveals new possibilities for controlling ruminant methane emissions. PLoS One 5(1): e8926
- Leahy *et al* (2013) The complete genome sequence of *Methanobrevibacter* sp. AbM4. Standards in Genomic Sciences 8(2): 215-227
- Lee *et al* (2013) Genome sequence of Methanobrevibacter sp. strain JH1, isolated from rumen of Korean native cattle. Genome Announcements 1(1): e00002-13
- Ley *et al* (2008) Worlds within worlds: evolution of the vertebrate gut microbiota. Nature Reviews Microbiology 6(10): 776-788
- Li (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint 1303:3997
- Mayer and Müller (2013) Adaptations of anaerobic archaea to life under extreme energy limitation. FEMS Microbiology Reviews 38: 449-472
- McMillan *et al* (2011) A_1A_0-ATP synthase of *Methanobrevibacter ruminantium* couples sodium ions for ATP synthesis under physiological conditions. Journal of Biological Chemistry 286(46): 39882-39892
- Miller and Lin (2002) Description of *Methanobrevibacter gottschalkii* sp. nov., *Methanobrevibacter thaueri* sp. nov., *Methanobrevibacter woesei* sp. nov. and *Methanobrevibacter wolinii* sp. nov. International Journal of Systematic and Evolutionary Microbiology 52: 819-822
- Patil *et al* (2012) The PhyloPythiaS web server for taxonomic assignment of metagenome sequences. PLoS One 7(6): e38581

- Paul *et al* (2012) “Methanoplasmatales,” Thermoplasmatales-related archaea in termite guts and other environments, are the seventh order of methanogens. Applied and Environmental Microbiology 78(23): 8425-8253
- Peng *et al* (2012) IDBA-UD: a de novo assembler for single cell and metagenomic sequence data with highly uneven depth. Bioinformatics 28(11): 1420-1428
- Piao *et al* (2013) Improved method for isolation of microbial RNA from biofuel feedstock for metatranscriptomics. Advances in Microbiology 3(1): 101-107
- Poulsen *et al* (2013) Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. Nature Communications 4:1428 doi:10.1038/ncomms2432
- Reeve *et al* (1997) Methanogenesis: genes, genomes and who’s on first? Journal of Bacteriology, 179(19): 5975-5986
- Reynolds *et al* (2008) Transmembrane topology and signal peptide prediction using dynamic Bayesian networks. PLoS Computational Biology 4(11): e1000213
- Samuel and Gordon (2006) A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. Proceedings of the National Academy of Sciences 103(26): 10011-10016
- Samuel *et al* (2007) Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. Proceedings of the National Academy of Sciences 104(25): 10643-10648
- Shi *et al* (2014) Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. Genome Research doi: 10.1101/gr.168245.113
- Reynolds *et al* (2008) Transmembrane topology and signal peptide prediction using dynamic Bayesian networks. PLoS Computational Biology 4(11): e1000213
- Tymensen *et al* (2012) Structures of free-living and protozoa-associated methanogen communities in the bovine rumen differ according to comparative analysis of 16S rRNA and mcrA genes. Microbiology 158(7): 1808-1817
- Van den Berg *et al* (2004) X-ray structure of a protein-conducting channel. Nature 427: 36-44
- Veenendaal *et al* (2004) The protein-conducting channel SecYEG. Biochimica et Biophysica Acta (BBA) - Molecular cell Research 1694(1-3): 81-95
- Wood *et al* (1982) Role of vitamin B_{12} in methyl transfer for methane biosynthesis by *Methanosarcina barkeri*. Science 216: 303-305
- Yao *et al* (2012) SVMTrip: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. PLoS One 7(9): e45152
- Yu and Morrison (2004) Improved extraction of PCR-quality community DNA from digesta and fecal samples. Biotechniques 36: 808-812

8. Appendices

Appendix 1. Growth medium used to determine the transcriptomic response of *M. ruminantium* M1 to increasing levels of selected metals *in vitro*

Component	Amount (per L)
Filter sterilised (0.22 µm) rumen fluid	25 ml
Salt solution A	50 ml
Salt solution B	50 ml
Metal solution 1	10 mL
Metal solution 2	0, 10 or 100 mL
Ammonium chloride	1 g
Sodium acetate	2.5 g
Sodium formate	2.5 g
Yeast extract	0.5 g
Tryptone	0.5 g
Resazurin (0.1% w/v stock)	1.0 ml
2-MES (10 mM stock)	1.0 mL
Sodium hydrogen carbonate	5.0 g
L-cysteine hydrochloride	0.5 g
Distilled water	to 1000 mL

Composition of salts solution A (20x stock)	Amount (per L)
NaCl	6.0 g
KH_2PO_4	3.0 g
$(NH_4)_2SO_4$	1.5 g
$CaCl_2 \cdot 2H_2O$	0.79 g
$MgSO_4 \cdot 7H_2O$	1.2 g
Distilled water	to 1000 ml

Composition of salts solution B (20x stock)	Amount (per L)
$K_2HPO_4 \cdot 3H_2O$	7.86 g
Distilled water	to 1000 ml

Composition and recipe for metal solution 1 (100x stock)

1. Dissolve 1.5 g nitrilotriacetic acid in 600mL dH_2O
2. Adjust pH to 6.5 with 3N KOH
3. Dissolve the following constituents in this order with constant mixing:

Component	Amount
$MgSO_4.7H_2O$	3.0 g
$MnSO_4.H_2O$	0.45 g
NaCl	1.0 g
$CaCl_2.2H_2O$	0.1 g
$CuSO_4.5H_2O$	0.01 g
$AIK(SO_4)_2.12H_2O$	0.018 g
H_3BO_3	0.01 g

4. Adjust pH to 7.0
5. Bring the final volume to 2 litres with dH_2O

Composition and recipe for metal solution 2 (100x stock)

1. Dissolve 1.5 g nitrilotriacetic acid in 600mL dH_2O
2. Adjust pH to 6.5 with 3N KOH
3. Dissolve the following constituents in this order with constant mixing:

Component	Amount
$FeSO_4.7H_2O$	0.1 g

$CoSO_4.7H_2O$	0.18 g
$ZnSO_4.7H_2O$	0.18 g
$Na_2MoO_4.2H_2O$	0.1 g
$NiSO_4.6H_2O$	0.1 g
Na_2SeO_4	0.19 g
$Na_2WO_2.2H_2O$	0.1 g

4. Adjust pH to 7.0
5. Bring the final volume to 2 litres with dH_2O

Appendix 2a. Description of rumen fluid samples provided by Kennedy and Charmley, 2012

Period	Description of diet	Samples
P1-P3	Low quality Black Speargrass	6
P1-P3	Low quality Black Speargrass + 20% Dolichos	5
P1-P3	Very low quality Black Speargrass	4
P1-P3	Very low quality Black Speargrass + 20% Dolichos	5
P1-P3	Very low quality Black Speargrass + 40% Dolichos	6
P1-P3	Very low quality Black Speargrass + urea/sulfur	5
P4	Medium quality Buffel grass	4
P4	Medium to high quality Buffel grass	4
P4	High quality Buffel grass	4
P6	Bisset creeping Bluegrass (26% leaf)	6
P6	Bisset creeping Bluegrass (41% leaf)	6
P8-P9	Rhodes grass	5
P8-P9	Rhodes grass + 20% Burgundy bean	5
P8-P9	Rhodes grass + 40% Burgundy bean	5
P8-P9	Rhodes grass + 20% Stylo	5
P8-P9	Rhodes grass + 40% Stylo	5

Appendix 2b. Description of rumen fluid samples provided by Li and Vercoe (unpublished)

The following paragraph has been taken from Xixi's thesis and included in this report to provide some background information about the samples.

Sheep (1-year-old Merino × Suffolk wethers) were randomly assigned to one of two diets (a control diet of oaten chaff/lupins mix or a test diet of oaten chaff/lupins and 15% E. glabra; n=10 individual animals per diet). The diets were fed for 30 days before feed digestibility, CH_4 output and rumen fermentation were measured. Compared to the control diets, feeding E. glabra reduced CH_4 emissions by 11.8% per gram of dry matter (DM) intake and 14.8% CH_4 per gram of digested DM, based on analysis of 10 animals per treatment group.

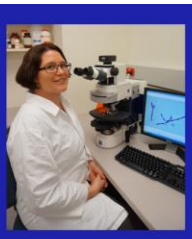
Appendix 3. Poster presented at GSC 17 and ASM 2015

Searching for methanogen genomic needles in a rumen microbiome haystack

Carly P. Rosewarne*, Mark I. Bradbury, Stuart E. Denman and Christopher S. McSweeney

* Fulbright Professional Scholar in Climate Change and Clean Energy
carly.rosewarne@csiro.au

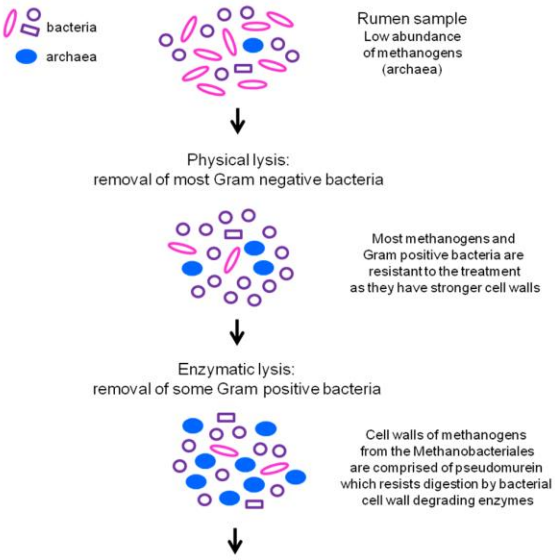
COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION, AUSTRALIA
www.csiro.au



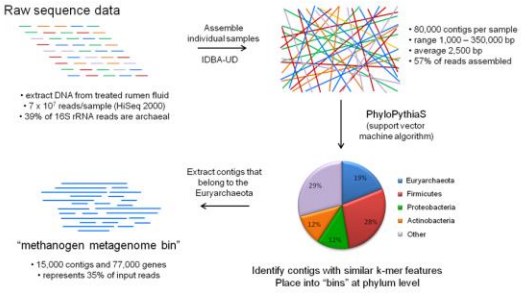
Many surveys continue to be undertaken to catalogue the key microbes responsible for methane production in livestock ("who's there?"). We are missing the important functional genomic information required to support development of targeted approaches for reducing methane emissions. Through the use of novel methods to study the genetic potential and gene expression profiles of rumen methanogens, it has been possible to reveal novel mechanisms used to enable their survival and persistence under *in vivo* conditions

Enriching for DNA from methanogens in rumen samples

- Despite the large volumes of methane eructated (burped) by ruminants, methanogens only represent around 1% of the total prokaryotic cells in the rumen microbiome
- Traditional shotgun metagenomic approaches cannot yet provide sufficient depth of coverage to enable characterisation of methanogen functional metabolic potential at a population scale
- The lack of data is rate-limiting to current efforts aimed at developing targeted approaches for direct inhibition of methanogens in ruminants
- We developed a method to enrich for key populations in rumen samples prior to DNA extraction by exploiting differences in cell wall structure between methanogens (archaea) and bacteria
- Up to 40% of 16S ribosomal RNA reads in the resulting metagenomes were assigned to members of the *Methanobrevibacter* and *Methanosphaera* genera, indicating that the method provides significant enrichment of the targeted methanogens



Production of "rumen methanogen metagenome bins"



Using the contigs as a template to detect gene expression

- We generated rumen methanogen metagenome bins from ten beef cattle typical of a northern Australian production system
- Each bin contained an average of 6 x 10^4 genes, with > 90% of each dataset comprised of contigs from the genus *Methanobrevibacter*
- The bins were used as scaffolds to identify highly expressed genes in paired RNA-seq datasets (HTSeq; 6x10^7 mRNA reads per sample)
- The highly expressed methanogen genes encode enzymes and transporters associated with the central methanogenesis pathway (including utilisation of formate as a substrate), systems for cobalt uptake and cobalamin biosynthesis (essential for the function of key methanogenic enzymes) and a range of adhesin-like proteins
- Adhesins are large cell-surface exposed proteins that enable attachment to surfaces or other cells. Cell-to-cell attachment is used to facilitate interspecies hydrogen transfer from bacteria to methanogens
- Up to 5% of the ORFs in each methanogen metagenome bin were annotated as adhesin-like proteins. Our novel approach has enabled us to identify and prioritise those that are highly expressed by rumen methanogens under *in vivo* conditions

Concluding remarks

- This research represents a significant advance in our current understanding of the functional genetic potential of rumen methanogens, providing a tangible link between "who's there?" and "what are they doing?"
- The methods are amenable to studies of selected archaeal and bacterial lineages in diverse environments, and we welcome interest from potential collaborators

BIOINFORMATIC TOOLS

• IDBA-UD: Peng *et al.*, 2012: doi 10.1093/bioinformatics/bts174
• PhyloPythiaS: Patil *et al.*, 2012: doi 10.1371/journal.pone.0038581
• HTSeq: Anders *et al.*, 2014: doi 10.1093/bioinformatics/btu638

ACKNOWLEDGEMENTS

This project is supported by funding from Meat & Livestock Australia and the Australian Government Department of Agriculture as part of the Carbon Farming Futures Filling the Research Gap Program

Carly is supported by the Australian-American Fulbright Commission during her sabbatical at the US DOE Joint Genome Institute in 2015