

Scotland's Rural College

Identification, comparison and validation of robust rumen microbial biomarkers for methane emissions using diverse *Bos taurus* breeds and basal diets

Auffret, MD; Stewart, R; Dewhurst, RJ; Duthie, C-A; Rooke, JA; Wallace, RJ; Freeman, TC; Snelling, TJ; Watson, M; Roehe, R

Published in:
Frontiers in Microbiology

DOI:
[10.3389/fmicb.2017.02642](https://doi.org/10.3389/fmicb.2017.02642)

First published: 09/01/2018

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Auffret, MD., Stewart, R., Dewhurst, RJ., Duthie, C-A., Rooke, JA., Wallace, RJ., Freeman, TC., Snelling, TJ., Watson, M., & Roehe, R. (2018). Identification, comparison and validation of robust rumen microbial biomarkers for methane emissions using diverse *Bos taurus* breeds and basal diets. *Frontiers in Microbiology*, 8, [2642]. <https://doi.org/10.3389/fmicb.2017.02642>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Identification, comparison and validation of robust rumen microbial biomarkers for methane emissions using diverse *Bos Taurus* breeds and basal diets

Marc D. Auffret^{1*}, Robert Stewart², Richard J. Dewhurst¹, Carol-Anne Duthie¹, John Rooke¹, Robert J. Wallace³, Tom C. Freeman², Timothy J. Snelling³, Mick Watson², Rainer Roehe¹

¹Scotland's Rural College, United Kingdom, ²Roslin Institute, University of Edinburgh, United Kingdom, ³Rowett Institute, University of Aberdeen, United Kingdom

Submitted to Journal:
Frontiers in Microbiology

Specialty Section:
Microbial Symbioses

Article type:
Original Research Article

Manuscript ID:
313122

Received on:
29 Sep 2017

Revised on:
10 Dec 2017

Frontiers website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

Conceptualization: MDA and RR
Formal analysis: MDA and MW
Original writing: MDA and RR
Review and Editing: MDA, RJD, C-AD, JAR, RJW, TCF, RS, MW and RR
"All authors read and approved the final manuscript."

Keywords

Rumen microbiome, Methane, biomarkers, Metagenomics, DIETS

Abstract

Word count: 323

Previous shotgun metagenomic analyses of ruminal digesta identified some microbial information that might be useful as biomarkers to select cattle that emit less methane (CH₄), which is a potent greenhouse gas. It is known that methane production (g/kgDMI) and to an extent the microbial community is heritable and therefore biomarkers can offer a method of selecting cattle for low methane emitting phenotypes. In this study a wider range of *Bos Taurus* cattle, varying in breed and diet, was investigated to determine microbial communities and genetic markers associated with high/low CH₄ emissions. Digesta samples were taken from 50 beef cattle, comprising four cattle breeds, receiving two basal diets containing different proportions of concentrate and also including feed additives (nitrate or lipid), that may influence methane emissions. A combination of partial least square analysis and network analysis enabled the identification of the most significant and robust biomarkers of CH₄ emissions (VIP>0.8) across diets and breeds when comparing all potential biomarkers together. Genes associated with the hydrogenotrophic methanogenesis pathway converting carbon dioxide to methane, provided the dominant biomarkers of CH₄ emissions and methanogens were the microbial populations most closely correlated with CH₄ emissions and identified by metagenomics. Moreover, these genes grouped together as confirmed by network analysis for each independent experiment and when combined. Finally, the genes involved in the methane synthesis pathway explained a higher proportion of variation in CH₄ emissions by PLS analysis compared to phylogenetic parameters or functional genes. These results confirmed the reproducibility of the analysis and the advantage to use these genes as robust biomarkers of CH₄ emissions. Volatile fatty acid concentrations and ratios were significantly correlated with CH₄, but these factors were not identified as robust enough for predictive purposes. Moreover, the methanotrophic *Methylomonas* genus was found to be negatively correlated with CH₄. Finally, this study confirmed the importance of using robust and applicable biomarkers from the microbiome as a proxy of CH₄ emissions across diverse production systems and environments.

Funding statement

The project was supported by grants from the Biotechnology and Biological Sciences Research Council (BBSRC BB/N01720X/1 and BB/N016742/1). The research is based on data from experiments funded by the Scottish Government as part of the 2011-2016 commission, Agriculture and Horticulture Development Board (AHDB) Beef & Lamb, Quality Meat Scotland (QMS) and Department for Environment Food & Rural Affairs (Defra).

Ethics statements

(Authors are required to state the ethical considerations of their study in the manuscript, including for cases where the study was exempt from ethical approval procedures)

Does the study presented in the manuscript involve human or animal subjects: Yes

Please provide the complete ethics statement for your manuscript. Note that the statement will be directly added to the manuscript file for peer-review, and should include the following information:

- Full name of the ethics committee that approved the study
- Consent procedure used for human participants or for animal owners
- Any additional considerations of the study in cases where vulnerable populations were involved, for example minors, persons with disabilities or endangered animal species

As per the Frontiers authors guidelines, you are required to use the following format for statements involving human subjects: This study was carried out in accordance with the recommendations of 'name of guidelines, name of committee' with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the 'name of committee'.

For statements involving animal subjects, please use:

This study was carried out in accordance with the recommendations of 'name of guidelines, name of committee'. The protocol was approved by the 'name of committee'.

If the study was exempt from one or more of the above requirements, please provide a statement with the reason for the exemption(s).

Ensure that your statement is phrased in a complete way, with clear and concise sentences.

This study was conducted at the Beef and Sheep Research Centre of Scotland's Rural College (SRUC, Edinburgh, UK). The experiment was approved by the Animal Experiment Committee of SRUC and was conducted in accordance with the requirements of the UK Animals (Scientific Procedures) Act 1986.

In review

1 **Identification, comparison and validation of robust rumen**
2 **microbial biomarkers for methane emissions using diverse *Bos***
3 ***Taurus* breeds and basal diets**

4
5 Marc D. Auffret*¹, Robert Stewart², Richard J. Dewhurst¹, Carol-Anne Duthie¹, John A.
6 Rooke¹, R. John Wallace³, Tom C. Freeman⁴, Timothy Snelling³, Mick Watson^{2,4}, Rainer
7 Roehe¹.

8
9 * Corresponding author

10 ¹ SRUC, Edinburgh, United Kingdom, ² Edinburgh Genomics, The Roslin Institute and
11 R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, ³ Rowett Institute,
12 University of Aberdeen, Aberdeen, United Kingdom, ⁴ Division of Genetics and Genomics,
13 The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, ⁴
14 Edinburgh Genomics, The Roslin Institute and R(D)SVS, University of Edinburgh,
15 Edinburgh, United Kingdom.

16
17 **Correspondence:**

18 **Marc.Auffret@sruc.ac.uk**

In review

19 Abstract

20 Previous shotgun metagenomic analyses of ruminal digesta identified some microbial
21 information that might be useful as biomarkers to select cattle that emit less methane (CH₄),
22 which is a potent greenhouse gas. It is known that methane production (g/kgDMI) and to an
23 extent the microbial community is heritable and therefore biomarkers can offer a method of
24 selecting cattle for low methane emitting phenotypes.

25 In this study a wider range of *Bos Taurus* cattle, varying in breed and diet, was investigated
26 to determine microbial communities and genetic markers associated with high/low CH₄
27 emissions. Digesta samples were taken from 50 beef cattle, comprising four cattle breeds,
28 receiving two basal diets containing different proportions of concentrate and also including
29 feed additives (nitrate or lipid), that may influence methane emissions. A combination of
30 partial least square analysis and network analysis enabled the identification of the most
31 significant and robust biomarkers of CH₄ emissions (VIP>0.8) across diets and breeds when
32 comparing all potential biomarkers together. Genes associated with the hydrogenotrophic
33 methanogenesis pathway converting carbon dioxide to methane, provided the dominant
34 biomarkers of CH₄ emissions and methanogens were the microbial populations most closely
35 correlated with CH₄ emissions and identified by metagenomics. Moreover, these genes
36 grouped together as confirmed by network analysis for each independent experiment and
37 when combined. Finally, the genes involved in the methane synthesis pathway explained a
38 higher proportion of variation in CH₄ emissions by PLS analysis compared to phylogenetic
39 parameters or functional genes. These results confirmed the reproducibility of the analysis
40 and the advantage to use these genes as robust biomarkers of CH₄ emissions.

41 Volatile fatty acid concentrations and ratios were significantly correlated with CH₄, but these
42 factors were not identified as robust enough for predictive purposes. Moreover, the
43 methanotrophic *Methylobacter* genus was found to be negatively correlated with CH₄.
44 Finally, this study confirmed the importance of using robust and applicable biomarkers from
45 the microbiome as a proxy of CH₄ emissions across diverse production systems and
46 environments.

47

48 **Keywords: Rumen microbiome, Methane, Biomarkers, Metagenomics, Diets.**

49

50 INTRODUCTION

51 Recent metagenomic analyses have highlighted the exciting opportunity that rumen microbial
52 biomarkers of methane (CH₄) emissions could enable the selection by breeding of cattle
53 which emit less CH₄ and ultimately may lower agricultural greenhouse gas (GHG) emissions.
54 Ross *et al.* (2013) highlighted that this approach may surpass current prediction accuracies
55 that are based on the host genome, especially for traits that are difficult to measure and
56 largely influenced by the gut microbiome. Methane has a large impact on global warming,
57 being 28-fold more potent as a GHG than carbon dioxide (CO₂) (IPCC, 2014). It is one of the
58 main anthropogenic sources (IPCC, 2014) and ruminants are major producers of CH₄,
59 accounting for 37% of total GHG from agriculture in the UK (Cottle *et al.*, 2011). Methane
60 results as an end product of anaerobic microbial fermentation in the rumen and it significant
61 negative economic and environmental impacts on animal production (Johnson and Johnson,
62 1995). A limited number of archaeal taxa within Euryarchaeota are methane producers and
63 the genes involved in this process are well characterised (Thauer *et al.*, 2008, Leahy *et al.*,
64 2010, Borrel *et al.*, 2013). The hydrogenotrophic pathway catalysing the conversion of CO₂
65 to methane is dominant in the rumen, and occurs in *Methanobrevibacter* spp. (Hook *et al.*,

2010, Danielsson *et al.*, 2017). However, methylotrophic methanogenesis also occurs in the *Methanomassilliicoccales* group (Li *et al.*, 2016), converting methylamine or methanol derived from digestion of feed constituents to methane (Poulsen *et al.*, 2013, Vanwonderghem *et al.*, 2017). In addition, more work is needed to identify the bacterial populations interacting with methanogens for H₂ or involved in different metabolic pathways associated with lactate or volatile fatty acids (VFA) including propionate, butyrate or acetate which are known to impact differently methane emissions (Moss *et al.*, 2000, Janssen *et al.*, 2010, Wanapat *et al.*, 2015, Kamke *et al.*, 2016). For example, *Megasphaera elsdenii* is the major rumen bacterium involved in the acrylate pathway converting lactate to propionate and, in the absence of lactate, producing acetate and butyrate but not propionate from glucose (Hino *et al.*, 1994, Russell and Wallace, 1997). Higher abundance of bacteria populations involved in propionate metabolism is associated with reduced methane emissions compared to acetate metabolism because more H₂ is utilised per mole VFA thus reducing availability for methane production (Janssen *et al.*, 2010, Wanapat *et al.*, 2015). Methanotrophic populations within both archaea and bacteria are known to metabolize methane as a carbon and energy source but the impact of such populations in the rumen seems likely to be minor (Parmar *et al.*, 2015, Wallace *et al.*, 2015).

Strategies to lower methane emissions in animal production are becoming an important field of research with the aims to enhance fermentation end-products that are useful to the host and reduce GHG emissions (Immig *et al.*, 1996, Knapp *et al.*, 2014). It is well known that diet has an impact on the microbial community composition and the genes carried by these populations (Rooke *et al.*, 2014, Henderson *et al.*, 2015). Diets with a higher content of concentrate (e.g. grain) compared to a forage diet (e.g. grass and silages) tend to produce lower methane emissions. For example, Giger-Reverdin & Sauvant (2000) observed that maximum methane emissions occurred between 30 and 40% of grain-based concentrate in the diet. Many feed additives have been explored for their impact on methane emissions. Addition of nitrate or polyunsaturated lipids (e.g. from rapeseed or linseed oil) to the diet showed promising results (Veneman *et al.*, 2015, Guyader *et al.*, 2016). The percentage of concentrate as constituent of the diet strongly affected this inhibitory effect (Duthie *et al.*, 2017). Mechanisms behind this effect are partly explained by the possible inhibition of H₂ producers in the presence of oil whilst nitrate is thought to act as a competitor with methanogens for H₂ and may also be toxic to methanogens (Guyader *et al.*, 2015). Besides the use of different diets or additives, recent research has identified links between the rumen microbiome and the host animal (Roehe *et al.*, 2016, Duthie *et al.*, 2017, Malmuthuge and Guan, 2017) and it has been established that host genetics influences methane emissions (Pinares-Patiño *et al.*, 2013, Herd *et al.*, 2014). The rumen microbiome may be the link between host genetics and methane emissions. Therefore, the impact of basal diets, additives and breeds on the microbiome may be considered and evaluated for the identification of robust biomarkers of CH₄ emissions.

Until now, proxies to predict methane emission phenotypes based on rumen samples including phylogenetic, genomic or metabolomic markers have not been considered to be robust and accurate, and are also expensive (Negussie *et al.*, 2017). This limitation has been partly attributed to the low number of ruminants studied for the identification and validation of biomarkers. There are inherent difficulties comparing the results of direct quantitation of methanogens using qPCR across different studies due to differences in sampling methods or primer target (McCartney *et al.*, 2013). Quantitative PCR has produced conflicting results when correlating absolute methanogen abundance with CH₄ emissions (Mosoni *et al.*, 2011, Morgavi *et al.*, 2012). However, a stronger correlation was obtained calculating relative

114 abundance between the Archaea and Bacteria abundance (A:B ratio) in rumen digesta
115 samples (Wallace *et al.*, 2014).

116 Reliable knowledge about the relationship between CH₄ emissions and both the microbiome
117 and the metabolites released is very important for improving the identification of biomarkers
118 (McCartney *et al.*, 2013; Ross *et al.*, 2013).

119 Metagenomics permits the identification of all genes comprising the microbiome and enables
120 taxonomic characterisation of the microbial population. Metagenomics has been confirmed
121 to be a powerful method for studying the rumen microbiome (Roehe *et al.*, 2016, Wallace *et*
122 *al.*, 2017). Roehe *et al.* (2016) identified 20 genes as biomarkers of methane emissions using
123 a combination of metagenomics and partial least square analyses. Moreover, the same authors
124 showed that these genes clustered together within a genetic network providing a proof of
125 principle about the feasibility of breeding selection by targeting these genes within the rumen
126 microbiome. These preliminary results were obtained on a limited number of beef cattle
127 (n=8) selected as extreme methane emitters (low or high) and fed with two basal diets (forage
128 or concentrate). Therefore the possibility to use a large scale method like metagenomics on a
129 set of data from different breeds of beef cattle fed different diets and coupled with VFA
130 monitoring is a great opportunity to identify and validate robust biomarkers of CH₄
131 emissions.

132 The aim of this study was (i) to evaluate the effect of two basal diets, and additives on the
133 rumen microbiome of a selection of four beef livestock breeds and to identify robust
134 biomarkers of CH₄ emissions associated with the microbiome or microbial activities, (ii) the
135 identification of robust biomarkers of CH₄ emissions associated with data from the microbial
136 community composition, the relative abundance of microbial populations, the relative
137 abundance of genes within the microbiome or VFA concentrations, all data collected from
138 three independent experiments and (iii) the comparison of these biomarkers to identify those
139 highly correlated with CH₄ emissions across diverse breeds and diets and the evaluation of
140 the possibility of implementing a breeding strategy using these microbial biomarkers from
141 the rumen microbiome.

142

143 **MATERIALS AND METHODS**

144 **Ethics statement**

145 This study was conducted at the Beef and Sheep Research Centre of Scotland's Rural College
146 (SRUC, Edinburgh, UK). The experiment was approved by the Animal Experiment
147 Committee of SRUC and was conducted in accordance with the requirements of the UK
148 Animals (Scientific Procedures) Act 1986.

149

150 **Animals, experimental design and diets**

151 In our previous study (Wallace *et al.*, 2015, Roehe *et al.*, 2016), data on feed efficiency and
152 methane emissions (measured using respiration chambers) were obtained from a 2 × 2
153 factorial design experiment of breed types and diets using 72 steers from a two-breed
154 rotational cross between Aberdeen Angus (AA) and Limousin (LIM) and completed in 2011.
155 Similar experiments were carried out using purebred Luing (LU) and crossbred Charolais
156 (CH) steers in 2013 and Aberdeen Angus (AA) and Limousin (LIM) rotational crossbred
157 steers in 2014. The data in this study were obtained from samples from those experiments
158 whereby animals with extreme high and low methane emissions (2011) or feed conversion
159 efficiency (2013 and 2014) were selected for whole genome sequencing. The breed type were

160 balanced within experiment comprising 4 AA and 4 LIM in 2011, 9 LU and 9 CH in 2013,
161 and 12 AA and 12 LIM in 2014. Methane emissions were measured individually for 48 h in
162 respiration chambers (Rooke *et al.*, 2014) and based on this result, 25 animals were
163 considered as low CH₄ emitters whilst the other 25 animals were classified as high CH₄
164 emitters. The average CH₄ emissions (g/kg DMI) between Low and High CH₄ emitters were
165 significantly different as shown in Fig 1. The animals were offered two complete diets *ad*
166 *libitum* consisting (g/kg DM) of approximately 500 forage to 500 concentrate or 80 forage to
167 920 concentrate which are subsequently referred to as forage and concentrate diets,
168 respectively. Nitrate, lipids or the combination of both were also added to the basal diet and
169 were compared with the control fed with the same diet without additive. The detailed diet
170 composition and proximate analysis has been reported previously by Rooke *et al.* (2014) and
171 Duthie *et al.* (2016, 2017). Animals were fed *ad libitum* during the entire experiment
172 including in the respiration chamber. A single sample of rumen fluid for VFA analysis
173 (expressed as molar proportions) was taken by stomach tube (naso ruminal sampling) within
174 1 h of cattle leaving the chambers in the 2011 experiment. VFA were determined in 2013 and
175 2014 using samples collected directly at the abattoir. As recommended by Terré *et al.* (2013),
176 we compared the VFA profiles between samples rather than total VFA concentrations
177 because of the different methods for rumen sampling applied. The acetate-to-propionate ratio
178 was calculated and considered as a proxy for H₂ generation.
179 Samples were obtained from a total of 50 animals balanced for breed type and diet and
180 including the eight post mortem samples previously studied in Roehe *et al.* (2016), (Table
181 S1).

182 **Genomic analysis**

183 The animals were fed *ad libitum* until they left the farm and thereafter slaughtered within
184 three hours in a commercial abattoir where two rumen fluid samples (approximately 50 mL)
185 were taken immediately after the rumen was opened to be drained. The main advantage to
186 collect rumen contents after slaughter is to obtain samples representative of both solid and
187 liquid phases. DNA was extracted from the rumen digesta samples following the protocol
188 described in Rooke *et al.*, (2014).

189 Illumina TruSeq libraries were prepared from genomic DNA and sequenced on an Illumina
190 HiSeq 2500 instrument (2011 samples) and on an Illumina HiSeq 4000 instrument (2013 and
191 2014 samples) by Edinburgh Genomics (Edinburgh, UK). Bioinformatics analyses using the
192 two sets of data followed the same procedure as previously described in Wallace *et al.*
193 (2015). Briefly, functional genes including the genes detailed in this study were identified
194 using KEGG genes database (<http://www.kegg.jp>). Genes with a relative abundance greater
195 than 0.001% were carried forward for downstream analysis.

196 For 16S rRNA gene analysis, the genomic reads were aligned to the Greengenes database
197 (DeSantis *et al.*, 2006) using Novoalign (www.novocraft.com) and also using the Kraken
198 database (Wood and Salzberg, 2014).

199 Parameters were adjusted such that all hits were reported that were equal in quality to the best
200 hit for each read, and allowing up to a 10 % mismatch across the fragment. Further details are
201 included in Wallace *et al.* (2015). These data can be downloaded from the European
202 Nucleotide Archive under accession PRJEB10338 and PRJEB21624.

203 **Statistical analysis**

204 Statistical analysis of the metagenomics samples was based on the complete sample profiles
205 as expressed by the pattern of metagenomic reads classified within KEGG orthologue groups
206 with >90% similarity and belonging to a single KEGG orthologue (KO) groups and the
207 relative abundance (percentage) of individual KO group in each profile. Principal coordinate
208
209

210 analysis (PCoA) was carried out using Gen-Stat 16th edition (VSN International Ltd, UK) to
211 identify the factors explaining differences observed in the microbial community (phylum
212 level) between samples. Relative abundance of microbial populations and functional genes,
213 Archaea-to-Bacteria (A:B) ratio, Firmicutes-to-Bacteroidetes (F:B) ratio as an indicator of
214 degradation activities carried by the two main phyla in rumen and acetate-to-propionate ratio
215 were compared using General Linear Models and *P*-values were Bonferroni corrected for
216 multiple testing (SPSS Statistics 22, IBM, USA).

217 In a network analysis using BioLayout Express3D (Freeman *et al.*, 2007), we identified the
218 distinct functional clusters of microbial genes for each experiment. These networks consist of
219 nodes representing microbial genes and the connecting edges determining the functional
220 linkages between these genes.

221 Partial least squares analysis (PLS, Version 9.1 for Windows, SAS Institute Inc., Cary, NC,
222 USA) was used to identify the most correlated microbial populations (at the phylum or genus
223 level) or microbial genes associated with methane emissions. This method was successfully
224 applied for the identification of microbial biomarkers in Wallace *et al.* (2015) and Roehe *et*
225 *al.* (2016). The PLS analysis accounted for multiple testing and the correlation between
226 microbial populations or genes as microbial parameters. In addition to microbial parameters,
227 the model included the diet effect (abiotic effector) and additionally the breed type effect
228 (host genetics effect). The model selection was based on the variable importance for
229 projection (VIP) criterion (Wold, 1995), whereby microbial parameters with a $VIP < 0.8$
230 contribute little to the prediction. Finally, a comparison between different factors identified as
231 highly correlated with CH₄ emissions and therefore considered as potential biomarkers were
232 tested by PLS analysis. In this study, biomarkers of CH₄ emissions will be considered as
233 robust when a similar result is observed across diverse diets and breeds and by comparing all
234 potential biomarker together. A robust biomarker may strengthen the confidence of
235 identifying low- versus high-emitting cattle. Those factors identified to be significant from
236 the microbial community composition, the relative abundance of microbial populations or
237 genes or VFA concentrations. All samples without VFA measurements were removed (N1,
238 N3, N7 and RR41).

239 The residual methane emissions were calculated using a General Linear Model including diet
240 and breed into the model and measured methane data as dependent variable. These residual
241 methane emissions are thus corrected for diet and breed and were centered and standardised
242 and only used when biomarkers were compared together.

243 Spearman's correlation analysis was also carried out to determine which factors (the same
244 factors tested by PLS) are correlated with CH₄ emissions using SPSS Statistics 22. *P*-
245 values ≤ 0.05 were considered significant and tendencies were represented (*P*-values < 0.1).

246

247 **RESULTS**

248 **Factors influencing the differences observed in methane emissions**

249 Several grouping conditions were tested using methane emission values from three
250 independent trials (Fig. 1). Average CH₄ emissions were 20.89 ± 0.75 g/kg dry matter intake
251 based on measurements from 50 animals. CH₄ emissions were 1.48-fold higher in the high-
252 CH₄ group ($P < 0.001$). CH₄ emissions were also higher in animals fed the forage compared to
253 concentrate basal diet ($P < 0.001$).

254 CH₄ emissions showed strong correlations with acetate ($F = 0.582$, $P < 0.001$), propionate ($F =$
255 0.574 , $P < 0.001$) and valerate ($F = 0.571$, $P < 0.001$) concentrations and to a lesser extent
256 isovalerate concentration ($F = 0.347$, $P < 0.05$) but not with butyrate or isobutyrate
257 concentrations (Table S1). Acetate-to-propionate ratio was strongly positively correlated

258 ($P<0.001$) with CH₄ emissions (Figure S1A). When samples were divided based on diet
259 treatment, this significant correlation disappeared in presence of concentrate (Fig S1B) and
260 only a tendency was found with the forage diet ($P=0.08$; Fig S1C).

261 **Change in microbial community composition between CH₄ emitters and diet treatments**

262 Using the Kraken database for the identification of the 16S rRNA sequences (Phylum level)
263 within the 50 metagenomics datasets, the difference observed within the microbial
264 community composition represented 36.9% over the first two principal coordinate analysis
265 axes (Fig. S2) and 45.2% when the third axis was included (data not shown).

266 The most abundant bacterial phyla (on average) identified were Firmicutes (42.8%),
267 Bacteroidetes (38.6%), Proteobacteria (6.6%), Fibrobacteres (4.9%) and Actinobacteria
268 (2.4%) representing on average 95.3% of the total community (Fig. S3). Proteobacteria was
269 the only dominant phylum significantly different with a higher abundance in low-CH₄
270 samples compared to high-CH₄ samples ($P=0.03$). Lower abundant phyla such as
271 *Deinococcus-Thermus* (0.12%, $P=0.006$), *Chlorobi* (0.07%, $P=0.003$), *Kiritimatiellaeota*
272 (0.01%, $P=0.02$), *Verrucomicrobia* (0.12%, $P=0.04$) and *Calditrichaeota* (0.003%, $P=0.01$)
273 were also identified as significantly different and generally with a higher abundance in high-
274 CH₄ samples except for *Calditrichaeota*. Comparing the effect of forage or concentrate diets,
275 a limited number of bacterial phyla ($n=4/32$) were affected, which were based on their
276 relative abundance in the rumen minor populations (Table S2). In general, the relative
277 abundance of microbial populations impacted by additives was higher in control treatment
278 except *Calditrichaeota* and *Proteobacteria*, the latter being 1.2-fold higher in presence of
279 nitrate compared to the control concentrate treatment (Table S2). On average, the Firmicutes-
280 to-Bacteroidetes ratio was at 1.22 and not significantly different between methane emitters or
281 diet treatments. *Euryarchaeota* were not impacted by nitrate or RSC in either concentrate or
282 forage diets.

283 The archaeal community represented $5.33 \pm 0.37\%$ of the total microbial community based
284 on 16S rRNA sequences and higher Shannon diversity was characterized using the Kraken
285 database compared to Greengenes as shown in Fig. 2A, with the former identifying more
286 methanogenic groups capable of utilizing acetoclastic, hydrogenotrophic and methylotrophic
287 pathways to produce methane (Fig. S4). The hydrogenotrophic pathway was highly
288 represented in the rumen content of both high- and low-methane emitting animals, mostly in
289 high emitters and represented on average 96.8% of total methanogens. The relative
290 abundance of total methanogens was double in high emitters compared to low emitters. This
291 result was explained by the significant dominance of several populations including
292 *Methanobrevibacter* (on average 94% of the methanogens), *Methanobacterium*,
293 *Methanococcus* and *Methanoculleus* species (Fig. 2A). On the other hand, the dominant
294 methylotrophic methanogen belonging to *Methanomassiliicoccales* order was identified as
295 *Candidatus Methanomethylophilus*, with a relative abundance 7-fold significantly higher in
296 the rumen microbiome of low-methane emitters compared to high-methane rumen samples
297 (Fig. 2A). Finally, the dominant acetoclastic methanogen was *Methanosarcina* species and
298 represented on average 0.4% of total methanogens (Fig. 2A). Overall, Shannon diversity
299 index calculated for total microbial community did not show any significant differences
300 between groups of methane emitters or diet. Focusing on methanogens, a higher diversity in
301 low emitters was confirmed with a Shannon diversity index of 0.55 compared to 0.28 in high
302 emitters ($P<0.001$). Effect of the additives on the relative abundance of methanogen
303 populations was not significant whilst methylotrophic methanogenic populations were on
304 average 2.27-fold more abundant in the concentrate diet supplemented with nitrate compared

305 to the control condition and only on average 1.21-fold higher in forage diet supplemented
306 with nitrate compared to the control treatment.

307 Using the Greengenes annotation, both methanogen diversity (Shannon index H) and
308 composition were lower and only represented by three dominant genera. However, the
309 general results on the dominant populations, methanogen diversity and the importance of
310 methylotrophic methanogens in low-methane emitters were the same but it has to be
311 considered that using this database the minor populations (e.g. acetoclastic methanogens)
312 were not recovered (Figs 2B and S4B).

313 Methanotrophic populations were also identified when using the Kraken database,
314 representing a limited part of the microbial community and being about 70-fold less abundant
315 than methanogens (on average $0.1 \pm 0.01\%$). This microbial group was highly dominated by
316 three methanotrophic bacteria including the genus *Methylobacterium* and to a lesser extent
317 *Methylomonas* and *Methylomicrobium* genera. However, only the *Methylomonas* genus was
318 different between emitters ($P=0.005$) or diet treatments ($P=0.005$) with a relative abundance
319 1.7-fold higher in low- compared to high-methane emitters. Finally, the diversity of
320 methanotrophic organisms was greater in high emitters ($P=0.02$) compared to low emitters
321 and there was no effect of diet or additives on methanotrophic populations.

322 **Identification of additional phylogenetic biomarkers of methane emissions**

323 The Archaea:Bacteria ratio was calculated for each sample and a positive correlation
324 ($P<0.001$) was confirmed by linear regression with methane emissions overall (Fig. 3). This
325 correlation was weaker when samples were grouped based on diet - being significant
326 ($P<0.01$) for the concentrate but not the forage diet (Fig. S5). Interestingly, a positive
327 correlation between CH_4 emissions and the relative abundance of Euryarchaeota was
328 confirmed ($F=0.567$, $P=0.003$) but only when studying high emitters.

329 Partial Least Square analysis including in the model diet and breed effects showed that the
330 relative abundances of 31 microbial genera were negatively correlated with methane
331 emissions (“Reducing effects on methane emissions” group in Table 1). There were 56
332 genera positively correlated (including 16 highly positively correlated) with methane
333 (“Increasing effects on methane emissions” group in Table 1) and 40 genera considered as
334 positively correlated with methane emissions but showing a low regression coefficients
335 (“Low effect on methane emissions” in Table 1). Moreover, the result generated by PLS and
336 including the 56 genera, breed type and diet effects, explained 50% of the variation in CH_4
337 emissions. One main result is that bacterial populations showed higher VIP value compared
338 to methanogens including the most abundant genus *Methanobrevibacter* and four other
339 hydrogenotrophic methanogens present at lower abundance including *Methanosphaera*
340 genus. Bacteria producing butyrate (e.g. *Butyrivibrio* and *Pseudobutyrvibrio* spp.) or CO_2
341 were positively correlated with CH_4 emissions, contrasting with those associated with amino
342 acid (e.g. *Acidaminococcus* and *Allisonella* species) and lactate metabolism (e.g.
343 *Megasphaera* and *Lactobacillus* genera) or populations consuming hydrogen (e.g.
344 *Dehalococcoides* genus). Other bacterial populations with significant VIP were known to be
345 associated with nitrogen (*Nitrosococcus* or *Nitrobacter* spp.) or sulphur cycles or those
346 classified in the average group were halotolerant populations or potentially involved in
347 organic matter breakdown, or syntrophic activities (e.g. *Syntrophobotulus* genus).

348 **Validation of functional genes as biomarkers of methane emissions**

349 The main result from the network analysis is that most of the same genes directly involved in
350 methane emissions were found over three independent trials and in one or two closed
351 clusters. For example, these genes grouped within a single cluster (C1) for the 2013 samples
352 or two clusters for the 2011 samples (C3 and C6) and 2014 samples (C3 and C5) (Figures
353 4ABC). Overall, 202 genes representing different microbial functions were identified using
354 KEGG in these clusters including those known to be involved in methane emissions (n=37).
355 However, only 27 genes associated with [high or low] methane emissions were detected in
356 the three experiments.

357 A PLS analysis using 202 genes (“general analysis”) was carried out and the results are
358 summarized in Table 2. As a result, 37 genes were identified as important to predict methane
359 emissions in cattle and as part of a model including breed type and diet effects explained 62%
360 of the variation in methane emissions. The most abundant of these were either subunits of the
361 methyl coenzyme M reductase gene catalysing the final step of CH₄ synthesis pathway
362 *mcrABG* (K00399, K00401 K00402) encoding for or genes associated with hydrogenase
363 activity, such as formate dehydrogenase, tetrahydromethanopterin S-methyltransferase,
364 formylmethanofuran dehydrogenase (K00123, K00125, K00577, K00580, K00581 and
365 K00584) or energy synthesis (V-type H⁺-transporting ATPase) (K02117 and K02118). The
366 former enzymes are associated with the hydrogenotrophic pathway while the genes encoding
367 for heterodisulfide reductase (K03389, K03390) and associated with low emitters, part of the
368 methylotrophic methanogenic pathway. All these genes were significantly higher in high-
369 emitting rumen samples compared to low-emitters ($P < 0.02$). Finally, the genes with a higher
370 VIP were not those encoding for the final reaction leading to CH₄ emissions but were
371 associated with the transfer of the methyl group (e.g. K06937) or hydrogen (e.g. K02117 and
372 K02118). In parallel, a similar PLS analysis was carried out but only using the genes (n=36)
373 known to be directly involved in the methane emissions pathway (Table S3). As a result, the
374 percentage of variation in methane emissions explained by these genes increased (65%)
375 compared to the general analysis (62%). Moreover, the genes with a higher VIP were not
376 those encoding for methyl-coenzyme M reductase (Table S3) as observed in the general
377 analysis.

378 **Comparison between the different biomarkers tested and correlation with CH₄** 379 **emissions**

380 Potential biomarkers were compared together by PLS analysis to evaluate the factors highly
381 correlated with CH₄ emissions (Table S4). Residual CH₄ emissions data were estimated to
382 remove the effect of diets and breeds and to allow the comparison of the potential biomarkers
383 identified by PLS as significantly correlated with CH₄ emissions. The PLS results identified
384 37 factors with a VIP value >0.80 and explaining 42% of the variation in residual CH₄ (Table
385 3). Within the 37 factors, 22 individual genes mostly involved in the hydrogenotrophic
386 methanogen pathway were identified. The other parameters identified included methanogen
387 populations (e.g. *Methanobrevibacter*, *Methanotorris* and *Methanohalophilus* genera), the
388 Shannon diversity indices for the methanogen community, PCoA scores or 6 bacterial
389 populations as well as the Archaea-to-Bacteria ratio. Finally, all the other parameters
390 previously tested and including the Acetate-to-Propionate ratio, or the data on the
391 methanotrophs (relative abundance) were not identified as final biomarkers. A different result
392 was obtained when a Spearman correlation test was applied on the same set of data using
393 non-corrected methane values and therefore still considering the effects of diet and breed
394 (Table S5). For example, Acetate:Propionate ratio showed the highest correlation with CH₄
395 emissions.

396

397 **DISCUSSION**

398 **Treatment effects on methane emissions**

399 In the present study, the results of three independent trials were compared and combined, and
400 the current analysis confirmed that the constituent of the basal diet was strongly and
401 significantly associated with CH₄ emissions. The proportion of dietary forage to concentrate
402 content in the diet as previously identified by Roehe *et al.* (2016) and Rooke *et al.* (2014).
403 The dietary additives used in this study as a strategy to lower CH₄ emissions did not show
404 significant results contrasting with previous works identifying nitrate and supplementary lipid
405 as some of the most promising methane mitigation additives in ruminants (Wallace *et al.*,
406 2014, Olijhoek *et al.*, 2015, Guyader *et al.*, 2016) while variations in response were detected
407 (Yang *et al.*, 2016). Factors that could explain these differences included the use of a reduced
408 number of rumen samples from animals initially selected for low- and high-feed conversion
409 and also the variability in the basal diet composition.

410 **Identification of functional genes as biomarkers of methane emissions**

411 In this combined analysis, most of the genes previously identified by Wallace *et al.* (2015)
412 and Roehe *et al.* (2016) were in general also identified in this study (n=19/20) by PLS
413 analysis over the 3 independent experiments and confirmed as strong biomarkers of CH₄
414 emissions. Most of these genes were involved in the hydrogenotrophic methane synthesis
415 pathway and grouped in one cluster or two attached clusters over the three independent
416 experiments as previously highlighted by Roehe *et al.* (2016). This study is one of the first
417 confirming the importance of genes encoding for heterodisulfide reductase in the rumen over
418 the genes associated with methylamine compounds or methanol conversion to accomplish the
419 first step of methylotrophic methanogenic pathway (Buan *et al.*, 2010, Borrel *et al.*, 2013).
420 Although genes encoding for heterodisulfide reductase were confirmed to be significantly
421 correlated with methane in the rumen of low emitters, the result of the biomarker comparison
422 did not identify those genes as robust biomarkers of CH₄ emissions. This result confirmed the
423 dominance of hydrogenotrophy over methylotrophy in the rumen (Hook *et al.*, 2010,
424 Danielsson *et al.*, 2017) but also highlighted that both pathways are important in explaining
425 methane emissions (Poulsen *et al.*, 2013). Interestingly, these genes associated with high VIP
426 value were in the upper part of the pathway and encode for methyltransferase, hydrogenase or
427 dehydrogenase activities but not directly the genes (e.g. *mcrA*) encoding for the methyl
428 coenzyme M reductase system the final step in methane production. This result tends to
429 confirm the importance of hydrogen concentration and thermodynamics affecting the
430 microbial communities and therefore VFA production and methane emissions (Wolin *et al.*,
431 1997, Rooke *et al.*, 2014). Contrasting with Shi *et al.* (2014), this study confirmed a
432 significant increase in the relative abundance of most of the genes involved in CH₄ emissions
433 by metagenomics, but in agreement with the same authors, not only *mcrA*, but all genes are
434 important in explaining higher CH₄ emissions. These results could explain weak correlations
435 previously observed with CH₄ emissions when targeting directly 16S rRNA gene or *mcrA*
436 (Morgavi *et al.*, 2012, Tapio *et al.*, 2017). There have been estimated unexpected negative
437 associations of microbial gene abundances and methane emissions (Table 2). Several reasons
438 including bioinformatics limitation (e.g. gene annotation error in database), the presence of
439 artefacts in the generated prediction model and a lack of biological knowledge for the genes
440 correlated with methane emissions results into the difficulty to associate estimates obtain

441 here with a mechanistic function. For example, it is known that different methanogen species
442 found in rumen samples carry most of the genes identified in this study. However, some
443 specific methanogenic species will lack a specific gene or a subunit within an operon as
444 described in Kaster *et al.* (2011). Therefore some species have a different impact on the
445 relative abundance of a specific subunit gene compare to others within the same operon and,
446 in consequence, on the coefficient value obtained by PLS analysis.

447 Although it would be of further interest to identify to which organisms these genes belong to,
448 this is beyond the scope of this paper and has to be addressed in substantial more detail using
449 different methodologies to provide accurate results. In addition, phylogenetic association with
450 the functional genes studied here is still challenging and was not carried out to avoid wrong
451 conclusions. This decision was made based on the fact that new methanogens are still
452 discovered (see Vanwonterghem *et al.*, 2016) and not necessarily carrying all the genes
453 involved in the methane synthesis pathway. Furthermore, different clades have been
454 identified and were even within the same genus (e.g. *Methanobrevibacter* SGMT or RO
455 clade) differently correlated with methane emissions in the same samples (Tapio *et al.*, 2017).
456 Specifically, *Methanobrevibacter* clade SGMT but not RO, was found more abundant in low
457 emitters while genera within methylophilic methanogens were enriched in high emitting
458 cattle..

459

460 **Most important phylogenetic parameters impacting on methane emissions**

461 Within the taxonomic parameters tested, factors directly associated with methanogens were
462 confirmed to be robust biomarkers, especially the relative abundance of *Methanobrevibacter*
463 genus. This genus is known to be the most dominant and active in the rumen (Henderson *et*
464 *al.*, 2015, Hook *et al.*, 2010, Tapio *et al.*, 2017, Wang *et al.*, 2017) and is also associated with
465 higher CH₄ emissions as confirmed here. Using the Kraken database, a wider diversity of
466 methanogens in the rumen was found compared to the results obtained using the Greengenes
467 database. This confirms preliminary observations by Poulsen *et al.* (2013) and Henderson *et*
468 *al.* (2015), and also highlights the importance of the reference database used to characterise
469 metagenomics data (Siegwald *et al.*, 2017).

470 Sun *et al.* (2012) confirmed that not all methanogens are active continuously in a
471 methanogenic environment and suggested that the availability of substrates was an important
472 cue for population growth. For example, *Methanocaldococcus* spp., *Methanotorris* spp. and
473 the methylophilic methanogen *Methanohalophilus* spp. were three low abundance genera
474 that were highly correlated with CH₄ emissions and identified as robust biomarkers across
475 different diets and breeds which contrasted with the result for the main methylophilic
476 methanogen *Candidatus Methanomethylophilus*. Moreover, the possibility to use these
477 biomarkers offers an efficient and cheaper alternative to metatranscriptomics considered as
478 more accurate tool to predict methane emissions compared to metagenomics (Shi *et al.*, 2014,
479 Wallace *et al.*, 2017). Finally, the identification of low abundance methanogen populations
480 but not all the most abundant as robust biomarkers may also explain weaker correlations
481 found between total methanogens and CH₄ emissions when 16S rRNA or *mcrA* genes were
482 targeted by qPCR (Mosoni *et al.*, 2011; Morgavi *et al.*, 2012). On the other hand, this weak
483 correlation can also be the result of methane oxidation by methanotrophs. This study is one of
484 the first confirming a greater abundance of methanotrophic populations, especially
485 *Methylophilus* genus in rumen and being significantly negatively correlated with CH₄
486 emissions. Genes associated with methanotrophy were not identified in this study and
487 previously in the set of eight animals (2011 experiment) as highlighted by Wallace *et al.*
488 (2015) and could be explained by not enough depth of sequencing for genes carried by very
489 low abundant populations (0.1%). The genus *Methylophilus* is identified in Greengenes and

490 Kraken databases but the last one contains a broader diversity of recently discovered
491 microbial populations that could improve the detection of low abundance genus in rumen
492 sample.

493 In terms of data directly associated with the microbial community composition, the
494 Archaea:Bacteria ratio was confirmed as a strong biomarker of CH₄ emissions while a lower
495 R-value (R=0.272) was found in this study compared to Wallace *et al.* paper (2014) which
496 calculated this ratio on a reduced number of cattle (R=0.49). This difference can be explained
497 by the initial set of 8 samples representing extreme methane emitters while the other 42
498 samples were not specifically selected for this trait. However, as also reported by the same
499 authors, this significant correlation was diet dependent, and was significant for concentrate
500 fed rumen samples but not forage samples. As previously shown in sheep by Kittelmann *et*
501 *al.* (2014), the microbial community composition (PCoA-2 in this study) even at the phylum
502 level was confirmed as robust biomarkers of CH₄ emissions. This could be explained by an
503 increase in the relative abundance of several bacterial populations within Firmicutes,
504 Bacteroidetes and Proteobacteria, mostly in low emitters as shown by the L/H ratio in Table
505 1. However, our study confirmed the necessity to calculate the methanogen diversity as
506 robust biomarker instead of total microbial diversity, not significantly different between
507 methane emitter groups in this study. These results differed from the idea developed by
508 Shabat *et al.* (2016) that cattle with higher CH₄ emissions will have higher total microbiome
509 diversity.

510 **Link between microbial communities and metabolites released**

511 In term of identifying links between the bacterial community containing most of the organic
512 matter degraders and the metabolites released in rumen, it seems that the degradation
513 activities carried out by the two most abundant bacterial phyla, Firmicutes and Bacteroidetes
514 as evaluated using the F:B ratio (Chen *et al.*, 2016) were not important to explain CH₄
515 emissions. More interestingly, this study confirmed the importance of other bacterial
516 populations associated with production of different metabolites which directly impacted on
517 CH₄ emissions and also showing a higher VIP value compared to methanogens (Table 2). For
518 example, *Butyrivibrio* spp. and *Pseudobutyrvibrio* spp. both butyrate-producing bacteria
519 were highly correlated with high CH₄ emissions while the presence of bacteria metabolizing
520 lactate (e.g. *Megasphaera*), degrading amino acids (e.g. *Acidaminococcus*) or competing for
521 H₂ were negatively correlated with CH₄ emissions (Kamke *et al.*, 2016, Sa *et al.*, 2016, Park
522 *et al.*, 2014). This result is explained by the different catabolic pathways carried by these
523 populations and directly impacting on H₂ partial pressure and subsequently on CH₄ emissions
524 (Kelly *et al.*, 2010, Janssen *et al.*, 2010, Kamke *et al.*, 2016, Sa *et al.*, 2016, Tapio *et al.*,
525 2017). The presence of lactate-utilising *Megasphaera* genus within the robust phylogenetic
526 biomarkers and negatively correlated with CH₄ emissions, highlighted the importance of
527 lactate metabolism controlling rumen fermentation (Counotte and Prins, 1981), production of
528 H₂ and specific VFAs and ultimately CH₄ (van Lingen *et al.*, 2016). The impact that VFAs
529 have on CH₄ emissions is established (Janssen *et al.*, 2010, Wanapat *et al.*, 2015) and the
530 positive correlation between different VFA or acetate-to-propionate ratio and CH₄ emissions
531 as previously stated by Shabat *et al.* (2016). However, none of the VFA factors were
532 identified as strong biomarkers (Table 3) confirming some contrasting results found between
533 VFA and CH₄ emissions and reviewed in Negussie *et al.* (2017). It could be explained by
534 necessity to study the relative inter-relationships among VFA measurements and also
535 between VFA and CH₄ yield as suggested by Palarea-Albaladejo *et al.* (2017). Therefore, the
536 impact that VFA have on CH₄ emissions may be less important compared to lactate

537 metabolism and new strategies for methane mitigation could be developed based on this
538 finding (Jeyanathan *et al.*, 2014).

539 Genera within *Succinivibrionaceae* known to be dominant in the digestive tract of the
540 Tammar wallaby, which emit one quarter of the methane emissions of the cattle (Pope *et al.*,
541 2011) were not identified within low emitters as previously shown by Wallace *et al.* (2015).
542 At the family level, the relative abundance of *Succinivibrionaceae* was 1.6 fold higher in low
543 CH₄ emitters (on average 1.3±0.2) compared to high emitters (on average 0.8±0.1) but
544 associated with a weak significance level ($P=0.049$). Surprisingly, these bacterial populations
545 were not identified as robust biomarkers, probably because of the functional redundancy
546 associated with the production or degradation of each metabolite. On the other hand, the
547 *Opitutus* genus was characterized as a robust biomarker and is known to be involved in H₂
548 production during the fermentation of organic matter (Chin *et al.*, 2001). Very little
549 information exists that explains the role of the *Dorea*, *Isosphaera*, *Faecalitalea*, *Colwellia*
550 and *Singulisphaera* on CH₄ emissions but some were associated with degradation capacities
551 in methane emitting environment (Kleindienst *et al.*, 2016).

552 Finally, we agree that other potential biomarkers of CH₄ emissions like archaeol could be
553 tested (McCartney *et al.*, 2013) and compared with the robust biomarkers identified in this
554 study. The same authors showed the benefit of using archaeol over qPCR method as a proxy
555 for CH₄ emissions.

556 To the best of our knowledge, this is the first report identifying and comparing potential CH₄
557 biomarkers across a range of dietary conditions and several experiments. This study confirms
558 the possible value of targeting functional genes using metagenomics as most of the robust
559 biomarkers identified were genes directly involved in the hydrogenotrophic methane
560 synthesis pathway while methylotrophic methanogens were also important in explaining CH₄
561 emissions. In addition, most of the genes directly involved in the methane synthesis pathway
562 grouped in the same cluster within a functional genes network and this result was reproduced
563 over three independent trials. Finally, this study confirm the significance of using robust and
564 applicable biomarkers from the microbiome as a proxy of CH₄ emissions across diverse beef
565 cattle breeds fed with different diets as an alternative for a trait that is difficult-to-measure on
566 a large number of animals. Moreover, the use of these biomarkers for the development of
567 molecular tools will help for the implementation of breeding strategies targeting low-methane
568 emitter animals.

569

570 **Conflict of Interest Statement**

571 This study was conducted in the absence of any commercial or financial relationships that could
572 be construed as a potential conflict of interest.

573 **Authors' contributions**

574 Conceptualization: MDA and RR

575 Formal analysis: MDA and MW

576 Original writing: MDA and RR

577 Review and Editing: MDA, RJD, C-AD, JAR, RJW, TCF, RS, MW and RR

578 "All authors read and approved the final manuscript."

579 **Funding**

580 The project was supported by grants from the Biotechnology and Biological Sciences
581 Research Council (BBSRC BB/N01720X/1 and BB/N016742/1). The research is based on

582 data from experiments funded by the Scottish Government as part of the 2011-2016
583 commission, Agriculture and Horticulture Development Board (AHDB) Beef & Lamb,
584 Quality Meat Scotland (QMS) and Department for Environment Food & Rural Affairs
585 (Defra).

586

587 **Acknowledgements**

588 We thank Irene Cabeza Luna, Andrew Southwell, Asier Zaragoza, Laura Nicoll, Lesley
589 Deans, and Claire Broadbent for the excellent technical support and Prof Alan Walker
590 (Rowett Institute, University of Aberdeen) for the critical reading of the manuscript.

591

592 **REFERENCES**

- 593 Borrel, G., O'Toole, P.W., Harris, H.M.B., Peyret, P., Brugère, J.-F., and Gribaldo, S. (2013).
594 Phylogenomic data support a seventh order of methylotrophic methanogens and
595 provide insights into the evolution of methanogenesis. *Genome Biol. Evol.* 5, 1769-
596 1780.
- 597 Buan, N.R., and Metcalf, W.W. (2010). Methanogenesis by *Methanosarcina acetivorans*
598 involves two structurally and functionally distinct classes of heterodisulfide reductase.
599 *Mol Microbiol.* 75, 843-853.
- 600 Chen, S., Cheng, H., Wyckoff, K.N., and He, Q. (2016). Linkages of Firmicutes and
601 Bacteroidetes populations to methanogenic process performance. *J Ind Microbiol*
602 *Biotechnol.* 43, 771-781.
- 603 Chin, K.-J., Liesack, W., and Janssen, P.H. (2001). *Opitutus terrae* gen. nov., sp. nov., to
604 accommodate novel strains of the division 'Verrucomicrobia' isolated from rice
605 paddy soil. *Int. J. Syst. Evol. Microbiol.* 51, 1965-1968.
- 606 Cottle, D.J., Nolan, J.V., and Wiedemann, S.G. (2011). Ruminant enteric methane mitigation:
607 a review. *Anim Prod Sci.* 51, 491-514.
- 608 Counotte, G.H.M., and Prins, R.A. (1981). Regulation of lactate metabolism in the rumen.
609 *Veterinary Research Communications* 5, 101-115.
- 610 Danielsson, R., Dicksved, J., Sun, L., Gonda, H., Müller, B., Schnürer, A., and Bertilsson, J.
611 (2017). Methane production in dairy cows correlates with rumen methanogenic and
612 bacterial community structure. *Front. Microbiol.* 8, 226.
- 613 Desantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., and Keller, K. (2006).
614 Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible
615 with ARB. *Appl Environ Microbiol.* 72, 5069-5072.
- 616 Duthie, C.-A., Haskell, M., Hyslop, J.J., Waterhouse, A., Wallace, R.J., Roehe, R., and
617 Rooke, J.A. (2017). The impact of divergent breed types and diets on methane
618 emissions, rumen characteristics and performance of finishing beef cattle. *Animal*, 1-
619 10.
- 620 Duthie, C.-A., Rooke, J.A., Troy, S., Hyslop, J.J., Ross, D.W., Waterhouse, A., and Roehe, R.
621 (2016). Impact of adding nitrate or increasing the lipid content of two contrasting
622 diets on blood methaemoglobin and performance of two breeds of finishing beef
623 steers. *Animal* 10, 786-795.
- 624 Freeman, T.C., Goldovsky, L., Brosch, M., Van Dongen, S., Mazière, P., and Grocock, R.J.
625 (2007). Construction, visualisation, and clustering of transcription networks from
626 microarray expression data. *PLoS Comput Biol.* 3, e206.
- 627 Giger-Reverdin, S., and Sauvant, D. (2000). Methane production in sheep in relation to
628 concentrate feed composition from bibliographic data. In: *Ledin, I.; Morand-Fehr, P.*

629 (Eds.) *Sheep and goat nutrition: Intake, digestion, quality of products and rangelands*
630 *Zaragoza. Zaragoza: Ciheam-Iamz, 2000. p.43-46.*

631 Guyader, J., Doreau, M., Morgavi, D.P., Gérard, C., Loncke, C., and Martin, C. (2016).
632 Long-term effect of linseed plus nitrate fed to dairy cows on enteric methane emission
633 and nitrate and nitrite residuals in milk. *Animal* 10, 1173-1181.

634 Guyader, J., Eugène, M., Meunier, B., Doreau, M., Morgavi, D.P., Silberberg, M., Rochette,
635 Y., Gérard, C., Loncke, C., and Martin, C. (2015). Additive methane-mitigating effect
636 between linseed oil and nitrate fed to cattle. *Journal of Animal Science* 93, 3564–
637 3577.

638 Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Collaborators, G., and Janssen,
639 P.H. (2015). Rumen microbial community composition varies with diet and host, but
640 a core microbiome is found across a wide geographical range. *Sci Rep.* 5, 14567.

641 Herd, R.M., Arthur, P.F., Bird, S.H., Donoghue, K.A., and Hegarty, R.S. (2014). Genetic
642 variation for methane traits in beef cattle. *Congress of Genetics Applied to Livestock*
643 *Production. Vancouver, Canada.*

644 Hino, T., Shimada, K., and Maruyama, T. (1994). Substrate preference in a strain of
645 *Megasphaera elsdenii*, a ruminal bacterium, and its implications in propionate
646 production and growth competition. *Applied Environmental Microbiology* 60, 1827-
647 1831.

648 Hook, S.E., Wright, A.-D.G., and McBride, B.W. (2010). Methanogens: methane producers of
649 the rumen and mitigation strategies. *Archaea* 2010, 945785.

650 Immig, I., Demeyer, D., Fiedler, D., Van Nevel, C., and Mbanzamihigo, L. (1996). Attempts
651 to induce reductive acetogenesis into a sheep rumen. *Arch. Tierernahr.* 49, 363-370.

652 Intergovernmental Panel on Climate Change (I.P.C.C.) (2014). Climate change 2014.
653 Synthesis report 2014.

654 Janssen, P.H. (2010). Influence of hydrogen on rumen methane formation and fermentation
655 balances through microbial growth kinetics and fermentation thermodynamics. *Anim*
656 *Feed Sci Technol.* 160, 1-22.

657 Jeyanathan, J., Martin, C., and Morgavi, D. (2014). The use of direct-fed microbials for
658 mitigation of ruminant methane emissions: a review. *Animal* 8, 250-261.

659 Johnson, K.A., and Johnson, D.E. (1995). Methane emissions from cattle. *J Anim Sci.* 73,
660 2483–2492.

661 Kamke, J., Kittelmann, S., Soni, P., Li, Y., Tavendale, M., Ganesh, S., Janssen, P.H., Shi, W.,
662 Froula, J., Rubin, E.M., and Attwood, G.T. (2016). Rumen metagenome and
663 metatranscriptome analyses of low methane yield sheep reveals a *Sharpea* enriched
664 microbiome characterised by lactic acid formation and utilisation. *Microbiome* 4, 56.

665 Kaster, A.K., Goenrich, M., Seedorf, H., Liesegang, H., Wollherr, A., Gottschalk, G., and
666 Thauer R.K. (2011). More than 200 genes required for methane formation from H₂
667 and CO₂ and energy conservation are present in *Methanothermobacter marburgensis*
668 and *Methanothermobacter thermotrophicus*. *Archaea.* 2011, 973848.

669 Kelly, W.J., Leahy, S.C., Altermann, E., Yeoman, C.J., Dunne, J.C., Kong, Z., Pacheco,
670 D.M., Li, D., Noel, S.J., Moon, C.D., Cookson, A.L., and Attwood, G.T. (2010). The
671 Glycobiome of the rumen bacterium *Butyrivibrio proteoclasticus* B316T highlights
672 adaptation to a polysaccharide-rich environment. *PLoS ONE.* 5, e11942.

673 Kittelmann, S., Pinares-Patiño, C.S., Seedorf, H., Kirk, M.R., Ganesh, S., Mcewan, J.C., and
674 Janssen, P.H. (2014). Two different bacterial community types are linked with the
675 low-methane emission trait in sheep. *PLoS One.* 9, e103171.

676 Kleindienst, S., Grim, S., Sogin, M., Bracco, A., Crespo-Medina, M., and Joye, S.B. (2016).
677 Diverse, rare microbial taxa responded to the Deepwater Horizon deep-sea
678 hydrocarbon plume. *ISME J.* 10, 400-415.

679 Knapp, J.R., Laur, G.L., Vadas, P.A., Weiss, W.P., and Tricarico, J.M. (2014). Invited
680 review: Enteric methane in dairy cattle production: Quantifying the opportunities and
681 impact of reducing emissions. *J. Dairy Sci.* 97, 3231-3261.

682 Leahy, S.C., Kelly, W.J., Altermann, E., Ronimus, R.S., Yeoman, C.J., Pacheco, D.M., Li,
683 D., Kong, Z., Mctavish, S., Sang, C., Lambie, S.C., Janssen, P.H., Dey, D., and
684 Attwood, G.T. (2010). The genome sequence of the rumen methanogen
685 *Methanobrevibacter ruminantium* reveals new possibilities for controlling ruminant
686 methane emissions. *PLoS ONE*, e8926.

687 Li, Y., Leahy, S.C., Jeyanathan, J., Henderson, G., Cox, F., Altermann, E., Kelly, W.J.,
688 Lambie, S.C., Janssen, P.H., Rakonjac, J., and Attwood, G.T. (2016). The complete
689 genome sequence of the methanogenic archaeon ISO4-H5 provides insights into the
690 methylotrophic lifestyle of a ruminal representative of the *Methanomassiliicoccales*.
691 *Standards in Genomic Sciences* 11, 59.

692 Malmuthuge, N., and Guan, L.L. (2017). Understanding host-microbial interactions in rumen:
693 searching the best opportunity for microbiota manipulation. *Microbiome* 8, 8.

694 McCartney, C.A., Bull, I.D., M., W.S., and Dewhurst, R.J. (2013). Technical note:
695 Comparison of biomarker and molecular biological methods for estimating
696 methanogen abundance. *J Anim Sci.* 91, 5724-5728.

697 Morgavi, D.P., Martin, C., Jouany, J.P., and Ranilla, M.J. (2012). Rumen protozoa and
698 methanogenesis: not a simple cause-effect relationship. *Br J Nutr.* 107, 388-397.

699 Mosoni, P., Martin, C., Forano, E., and Morgavi, D.P. (2011). Long-term defaunation
700 increases the abundance of cellulolytic ruminococci and methanogens but does not
701 affect the bacterial and methanogen diversity in the rumen of sheep. *J Anim Sci.* 89,
702 783-791.

703 Moss, A., Jouany, J.-P., and Newbold, J. (2000). Methane production by ruminants: its
704 contribution to global warming. *Ann. Zootech.* 49, 231-253.

705 Negussie, E., De Haas, Y., Dehareng, F., Dewhurst, R.J., Dijkstra, J., Gengler, N., Morgavi,
706 D.P., Soyeurt, H., Van Gastelen, S., Yan, T., and Biscarini, F. (2017). Invited review:
707 Large-scale indirect measurements for enteric methane emissions in dairy cattle: A
708 review of proxies and their potential for use in management and breeding decisions. *J.*
709 *Dairy Sci.* 100, 2433-2453.

710 Olijhoek, D.W., Hellwing, A.L.F., Brask, M., Weisbjerg, M.R., Højberg, O., Larsen, M.K.,
711 Dijkstra, J., J., E.E., and Lund, P. (2015). Effect of dietary nitrate level on enteric
712 methane production, hydrogen emission, rumen fermentation, and nutrient
713 digestibility in dairy cows. *J. Dairy Sci.* 99, 6191–6205.

714 Palarea-Albaladejo, J., Rooke, J.A., Nevison, I.M., and Dewhurst, R.J. (2017). Compositional
715 mixed modeling of methane emissions and ruminal volatile fatty acids from
716 individual cattle and multiple experiments. *J Anim Sci.* 95, 2467-2480.

717 Park, J., Park, S., and Kim, M. (2014). Anaerobic degradation of amino acids generated from
718 the hydrolysis of sewage sludge. *Environ Technol.* 35, 1133-1139.

719 Parmar, N.R., Nirmal Kumar, J.I., and Joshi, C.G. (2015). Exploring dietdependent shifts in
720 methanogen and methanotroph diversity in the rumen of Mehsani buffalo by a
721 metagenomics approach. *Frontiers in Life Science* 8, 371-378.

722 Pinares-Patiño, C.S., Hickey, S.M., Young, E.A., Dodds, K.G., Maclean, S., Molano, G.,
723 Sandoval, E., Kjestrup, H., Harland, R., Pickering, N.K., and Mcewan, J.C. (2013).
724 Heritability estimates of methane emissions from sheep. *Animal* 7, 316–321.

725 Pope, P.B., Smith, W., Denman, S.E., Tringe, S.G., Barry, K., Hugenholtz, P., Mcsweeney,
726 C.S., Mchardy, A.C., and Morrison, M. (2011). Isolation of *Succinivibrionaceae*
727 implicated in low methane emissions from Tammar wallabies. *Science* 333, 646-648.

- 728 Poulsen, M., Schwab, C., Jensen, B.B., Engberg, R.M., Spang, A., Canibe, N., Højberg, O.,
729 Milinovich, G., Fragner, L., Schleper, C., Weckwerth, W., Lund, P., Schramm, A.,
730 and Urich, T. (2013). Methylophilic methanogenic Thermoplasmata implicated in
731 reduced methane emissions from bovine rumen. *Nature Commun.* 4, 1428.
- 732 Roehe, R., Dewhurst, R.J., Duthie, C.-A., Rooke, J.A., Mckain, N., Ross, D.W., Hyslop, J.J.,
733 Waterhouse, A., Freeman, T.C., Watson, M., and Wallace, R.J. (2016). Bovine host
734 genetic variation influences rumen microbial methane production with best selection
735 criterion for low methane emitting and efficiently feed converting hosts based on
736 metagenomic gene abundance. *PLoS Genet.* 12, e1005846.
- 737 Rooke, J.A., Wallace, R.J., Duthie, C.-A., Mckain, N., De Souza, S.M., Hyslop, J.J., Ross,
738 D.W., Waterhouse, A., and Roehe, R. (2014). Hydrogen and methane emissions from
739 beef cattle and their rumen microbial community vary with diet, time after feeding
740 and genotype. *Br J Nutr.* 112, 398-407.
- 741 Ross, E.M., Moate, P.J., Maret, L., Cocks, B.G., and Hayes, B.J. (2013). Metagenomic
742 predictions: from microbiome to complex health and environmental phenotypes in
743 humans and cattle. *PLoS ONE* 8, e73056.
- 744 Russell, J.B., and Wallace, R.J. (1997). Energy-yielding and energy-consuming reactions. *In:*
745 *The Rumen Microbial Ecosystem. 2nd ed. (Ed. P. J. Hobson and C. S. Stewart),*
746 *Blackie Acad. Profess. London. pp. 246-282.*
- 747 Sa, J.H., Kwak, G.H., Han, K., Ahn, D., Cho, S.J., Lee, J.D., and Lee, K.H. (2016). Inhibition
748 of methane and natural gas hydrate formation by altering the structure of water with
749 amino acids. *Sci Rep.* 6.
- 750 Shabat, S.K., Sasson, G., Doron-Faigenboim, A., Durman, T., Yaacoby, S., Berg Miller,
751 M.E., White, B.A., Shterzer, N., and Mizrahi, I. (2016). Specific microbiome-
752 dependent mechanisms underlie the energy harvest efficiency of ruminants. *ISME J.*
753 10, 2958-2972.
- 754 Shi, W., Moon, C.D., Leahy, S.C., Kang, D., Froula, J., Kittelmann, S., Fan, C., Deutsch, S.,
755 Gagic, D., Seedorf, H., Kelly, W.J., Atua, R., Sang, C., Soni, P., Li, D., Pinares-
756 Patiño, C.S., Mcewan, J.C., Janssen, P.H., Chen, F., Visel, A., Wang, Z., Attwood,
757 G.T., and Rubin, E.M. (2014). Methane yield phenotypes linked to differential gene
758 expression in the sheep rumen microbiome. *Genome Research* 24, 1517-1525.
- 759 Siegwald, L., Touzet, H., Lemoine, Y., Hot, D., Audebert, C., and Caboche, S. (2017).
760 Assessment of common and emerging bioinformatics pipelines for targeted
761 metagenomics. *PLoS ONE* 12, e0169563.
- 762 Sun, C.L., Brauer, S.L., Cadillo-Quiroz, H., Zinder, S.H., and Yavitt, J.B. (2012). Seasonal
763 changes in methanogenesis and methanogenic community in three peatlands, New
764 York State. *Front. Microbiol.* 3, 81.
- 765 Tapio, I., Snelling, T.J., Strozzi, F., and Wallace, R.J. (2017). The ruminal microbiome
766 associated with methane emissions from ruminant livestock. *Journal of Animal*
767 *Science and Biotechnology* 8, 7.
- 768 Terré, M., Castells, L., and Fàbregas, F. (2013). Short communication: Comparison of pH,
769 volatile fatty acids, and microbiome of rumen samples from preweaned calves
770 obtained via cannula or stomach tube. *J. Dairy Sci.* 96, 5290-5294.
- 771 Thauer, R.K., Kaster, A.-K., Seedorf, H., Buckel, W., and Hedderich, R. (2008).
772 Methanogenic archaea: ecologically relevant differences in energy conservation.
773 *Nature Reviews Microbiology* 6, 579-591.
- 774 Van Lingen, H.J., Plugge, C.M., Fadel, J.G., Kebreab, E., Bannink, A., and Dijkstra, J.
775 (2016). Thermodynamic driving force of hydrogen on rumen microbial metabolism: a
776 theoretical investigation. *PLoS ONE* 11, e0161362.

- 777 Vanwonterghem, I., Evans, P.N., Parks, D.H., Jensen, P.D., Woodcroft, B.J., Hugenholtz, P.,
778 and Tyson, G.W. (2017). Methylophilic methanogenesis discovered in the archaeal
779 phylum Verstraetearchaeota. *Nature Microbiology* 1, 16170.
- 780 Veneman, J.B., Muetzel, S., Hart, K.J., Faulkner, C.L., Moorby, J.M., Perdok, H.B., and J.,
781 N.C. (2015). Does dietary mitigation of enteric methane production affect rumen
782 function and animal productivity in dairy cows? *PLoS ONE* 10, e0140282.
- 783 Wallace, R.J., Rooke, J.A., Duthie, C.-A., Hyslop, J.J., Ross, D.W., Mckain, N., Motta De
784 Souza, S., Snelling, T.J., Waterhouse, A., and Roehe, R. (2014). Archaeal abundance
785 in post-mortem ruminal digesta may help predict methane emissions from beef cattle.
786 *Sci Rep.* 4, 5892.
- 787 Wallace, R.J., Rooke, J.A., Mckain, N., Duthie, C.-A., Hyslop, J.J., Ross, D.W., Waterhouse,
788 A., Watson, M., and Roehe, R. (2015). The rumen microbial metagenome associated
789 with high methane production in cattle. *BMC Genomics* 16, 839.
- 790 Wallace, R.J., Snelling, T.J., McCartney, C.A., Tapio, I., and Strozzi, F. (2017). Application
791 of meta-omics techniques to understand greenhouse gas emissions originating from
792 ruminal metabolism. *Genet Sel Evol.* 16, 9.
- 793 Wanapat, M., Cherdthong, A., Phesatcha, K., and Kang, S. (2015). Dietary sources and their
794 effects on animal production and environmental sustainability. *Animal Nutrition* 1,
795 96-103.
- 796 Wang, Z., Elekwachi, C.O., Jiao, J., Wang, M., Tang, S., Zhou, C., Tan, Z., and Forster, R.J.
797 (2017). Investigation and manipulation of metabolically active methanogen
798 community composition during rumen development in black goats. *Sci Rep.* 7, 442.
- 799 Wold, S. (1995). PLS for multivariate linear modelling. *Van de Waterbeemd H, editor*
800 *Chemometric methods in molecular design VCH Publishers, Weinheim, Germany.*,
801 195–218.
- 802 Wolin, M.J., Miller, T.L., and Stewart, C.S. (1997). Microbe–microbe interactions. *In The*
803 *Rumen Microbial Ecosystem, pp. 467–491 [PN Hobson and CS Stewart, editors].*
804 *London: Chapman and Hall.*
- 805 Wood, D.E., and Salzberg, S.L. (2014). Kraken: ultrafast metagenomic sequence
806 classification using exact alignments. *Genome Biol. Evol.* 15, R46.
- 807 Yang, C., Rooke, J.A., Cabeza, I., and Wallace, R.J. (2016). Nitrate and inhibition of ruminal
808 methanogenesis: microbial ecology, obstacles, and opportunities for lowering
809 methane emissions from ruminant livestock. *Front. Microbiol.* 7, 132.

810 **Figures and tables**

811 Figure 1: Boxplots representing methane emissions under different conditions. High: High
812 methane emitters (n=25), Low: Low methane emitters (n=25), FOR: Forage (n=34), CONC:
813 Concentrate (n=16), CONT: all controls (n=20), NIT: all samples with nitrate (n=12), RSC:
814 all samples with supplementary lipid (n=12), Comb: all samples with nitrate and
815 supplementary lipid (n=6), AAX: all samples from Aberdeen Angus (n=13), CHx: all samples
816 from Charolais (n=12), LIMx: all samples from Limousin (n=13), LuIng: all samples from
817 LuIng (n=12). ** $P < 0.01$

818 Figure 2: Diversity of methanogen genera using (A) Kraken database or (B) Greengenes
819 database.

820 A: Acetoclastic methanogens, H: Hydrogenotrophic methanogens, M: Methylophilic
821 methanogens. ** $P < 0.01$, * $P < 0.05$ indicates different between low and high emitting groups.

822

823 Figure 3: Linear regression between Archaea:Bacteria ratio and CH₄ emissions.

824 Black circle: all samples. Equation for the linear regression was included in figure when the
825 difference was significant ($P < 0.05$).

826 Figure 4: Functional clusters of microbial genes identified using network analysis for (A) the
827 2011 experiment (n=1424 genes), (B) the 2013 experiment (n=1178 genes) (C) the 2014
828 experiment (n=1224 genes). Correlation analysis of microbial gene abundance was used to
829 construct networks, where nodes represent microbial genes and edges the correlation in their
830 abundance.

831 Table 1: PLS results identifying the most important microbial genera affecting methane
 832 emissions.

Phylum	Microbial Genus	VIP	Coef.	Mean Low CH4	Mean High CH4	L/H CH4 ratio	Function
Reducing effects on methane emissions							
Chloroflexi	<i>Dehalococcoides</i>	1.42	-0.039	0.044	0.020	2.15	H ₂ ox.
Bacteroidetes	<i>Odoribacter</i>	1.39	-0.040	0.058	0.028	2.09	commensal
Firmicutes	<i>Megasphaera</i>	1.35	-0.037	0.218	0.047	4.68	Lactate
Firmicutes	<i>Acidaminococcus</i>	1.34	-0.035	1.143	0.099	11.57	AA
Firmicutes	<i>Jeotgalicoccus</i>	1.23	-0.035	0.003	0.002	1.17	halotolerant
Firmicutes	<i>Allisonella</i>	1.16	-0.026	0.057	0.005	11.17	AA
Firmicutes	<i>Salinicoccus</i>	1.11	-0.028	0.004	0.003	1.57	halotolerant
Thermotogae	<i>Kosmotoga</i>	1.09	-0.028	0.003	0.002	1.31	thermophile
Bacteroidetes	<i>Mitsuokella</i>	1.08	-0.022	0.533	0.070	7.59	Phytate
Actinobacteria	<i>Olsenella</i>	1.07	-0.029	2.151	0.983	2.19	Lactate
Bacteroidetes	<i>Bacteroides</i>	1.02	-0.027	1.697	1.102	1.54	VFA
Firmicutes	<i>Dorea</i>	0.98	-0.024	0.106	0.062	1.70	Acetogen
Proteobacteria	<i>Wenzhouxiangella</i>	0.98	-0.001	0.016	0.005	3.28	halotolerant
Firmicutes	<i>Roseburia</i>	0.96	-0.022	0.172	0.079	2.18	Butyrate
Proteobacteria	<i>Edwardsiella</i>	0.96	-0.003	0.032	0.019	1.71	N.I.
Firmicutes	<i>Aneurinibacillus</i>	0.96	-0.026	0.005	0.003	1.66	Lignin degrader
Firmicutes	<i>Pelosinus</i>	0.96	-0.023	0.017	0.011	1.48	degrader
Proteobacteria	<i>Methylomonas</i>	0.95	-0.002	0.018	0.011	1.73	Methanotrophy
Firmicutes	<i>Veillonella</i>	0.94	-0.019	0.008	0.003	2.57	Lactate
Proteobacteria	<i>Halotalea</i>	0.94	-0.003	0.009	0.005	1.83	halotolerant
Proteobacteria	<i>Alkalilimnicola</i>	0.92	-0.001	0.010	0.007	1.41	halotolerant
Proteobacteria	<i>Sulfurovum</i>	0.92	-0.017	0.007	0.004	1.46	H ₂ ox.
Proteobacteria	<i>Colwellia</i>	0.92	-0.016	0.007	0.003	2.25	alkane degrader
Proteobacteria	<i>Marinomonas</i>	0.91	-0.016	0.004	0.003	1.39	halotolerant
Proteobacteria	<i>Nitrobacter</i>	0.90	0.000	0.009	0.006	1.42	NOB
Proteobacteria	<i>Thalassospira</i>	0.90	-0.023	0.003	0.003	1.16	halotolerant
Firmicutes	<i>Faecalitalea</i>	0.87	-0.024	0.020	0.010	2.02	AA
Euryarchaeota	<i>Methanohalophilus</i>	0.85	-0.022	0.002	0.001	1.50	Methanogen (M)
Firmicutes	<i>Lactobacillus</i>	0.83	-0.019	0.338	0.199	1.70	Lactate
Bacteroidetes	<i>Zobellia</i>	0.83	-0.021	0.003	0.002	1.27	mesophile
Proteobacteria	<i>Nitrosococcus</i>	0.83	-0.022	0.003	0.002	1.26	AOB
Low effect on methane emissions							
Actinobacteria	<i>Sanguibacter</i>	1.11	0.012	0.011	0.006	1.72	in blood
Proteobacteria	<i>Aromatoleum</i>	1.06	0.016	0.008	0.007	1.26	degrader
Proteobacteria	<i>Thiocystis</i>	1.04	0.009	0.014	0.010	1.46	sulphur
Proteobacteria	<i>Microbulbifer</i>	1.02	0.005	0.022	0.010	2.17	halotolerant

Euryarchaeota	<i>Halosimplex</i>	1.02	0.010	0.004	0.002	1.81	halotolerant
Proteobacteria	<i>Cronobacter</i>	1.00	0.002	0.042	0.020	2.08	pathogen
Actinobacteria	<i>Modestobacter</i>	0.99	0.004	0.010	0.005	2.19	halotolerant
Proteobacteria	<i>Neorickettsia</i>	0.99	0.011	0.002	0.001	1.35	pathogen
Proteobacteria	<i>Halorhodospira</i>	0.98	0.004	0.014	0.008	1.77	halotolerant
Proteobacteria	<i>Serratia</i>	0.98	0.008	0.061	0.048	1.27	N.I.
Spirochaete	<i>Salinispira</i>	0.98	0.010	0.008	0.005	1.39	halotolerant
Proteobacteria	<i>Asticcacaulis</i>	0.98	0.007	0.008	0.005	1.59	N.I.
Proteobacteria	<i>Sideroxydans</i>	0.98	0.003	0.012	0.005	2.29	Iron ox.
Proteobacteria	<i>Pantoea</i>	0.97	0.001	0.043	0.023	1.83	N.I.
Proteobacteria	<i>Agrobacterium</i>	0.97	0.007	0.041	0.032	1.29	N.I.
Proteobacteria	<i>Raoultella</i>	0.97	0.009	0.012	0.009	1.40	pathogen
Proteobacteria	<i>Halomonas</i>	0.96	0.002	0.043	0.027	1.59	halotolerant
Armatimonade tes	<i>Chthonomonas</i>	0.96	0.002	0.004	0.002	1.82	N.I.
Proteobacteria	<i>Ferrimonas</i>	0.96	0.005	0.010	0.006	1.59	Iron
Proteobacteria	<i>Acidihalobacter</i>	0.96	0.011	0.017	0.013	1.26	halotolerant
Actinobacteria	<i>Dermabacter</i>	0.95	0.003	0.007	0.004	1.95	N.I.
Proteobacteria	<i>Dokdonella</i>	0.95	0.001	0.011	0.007	1.64	N.I.
Proteobacteria	<i>Enterobacter</i>	0.95	0.004	0.061	0.046	1.32	degrader
Actinobacteria	<i>Tsukamurella</i>	0.95	0.002	0.007	0.004	1.90	degrader
Proteobacteria	<i>Immundisolibacter</i>	0.95	0.001	0.016	0.009	1.89	degrader
Proteobacteria	<i>Mesorhizobium</i>	0.95	0.007	0.055	0.042	1.31	degrader
Proteobacteria	<i>Lacimicrobium</i>	0.94	0.005	0.004	0.003	1.51	halotolerant
Proteobacteria	<i>Castellaniella</i>	0.94	0.008	0.014	0.010	1.33	N.I.
Proteobacteria	<i>Pseudomonas</i>	0.94	0.001	0.482	0.342	1.41	degrader
Proteobacteria	<i>Defluviimonas</i>	0.94	0.003	0.008	0.005	1.49	halotolerant
Dienococcus- Thermus	<i>Truepera</i>	0.93	0.007	0.009	0.007	1.28	degrader
Proteobacteria	<i>Methyloceanibacter</i>	0.93	0.003	0.008	0.006	1.40	Methylotrophy
Proteobacteria	<i>Thioflavicoccus</i>	0.93	0.004	0.015	0.011	1.40	sulphur
Firmicutes	<i>Syntrophobotulus</i>	0.92	0.003	0.008	0.005	1.49	Syntrophy
Proteobacteria	<i>Dyella</i>	0.91	0.004	0.028	0.021	1.33	degrader
Chlorobi	<i>Chlorobium</i>	0.90	0.002	0.021	0.016	1.32	sulphur
Cyanobacteria	<i>Microcoleus</i>	0.90	0.004	0.002	0.002	1.33	sulphur
Proteobacteria	<i>Chelativorans</i>	0.89	0.006	0.010	0.007	1.29	degrader
Proteobacteria	<i>Halioglobus</i>	0.89	0.002	0.005	0.004	1.36	halotolerant
Proteobacteria	<i>Pluralibacter</i>	0.88	0.003	0.014	0.010	1.33	pathogen
Increasing effects on methane emissions							
Proteobacteria	<i>Sedimenticola</i>	1.36	0.038	0.008	0.005	1.40	SOB
Firmicutes	<i>Sarcina</i>	1.33	0.038	1.142	3.246	0.35	CO ₂ prod.
Firmicutes	<i>Butyrivibrio</i>	1.31	0.037	2.107	3.017	0.70	Butyrate
Euryarchaeota	<i>Methanotorris</i>	1.30	0.036	0.002	0.003	0.58	Methanogen (H)
Euryarchaeota	<i>Methanobrevibacter</i>	1.23	0.034	4.166	7.146	0.58	Methanogen (H)
Planctomycete	<i>Isosphaera</i>	1.15	0.032	0.003	0.005	0.63	degrader

s							
Firmicutes	<i>Pseudobutyrvibrio</i>	1.13	0.032	0.434	0.617	0.70	Butyrate
Euryarchaeota	<i>Methanobacterium</i>	1.11	0.032	0.040	0.053	0.76	Methanogen (H)
Planctomycetes	<i>Singulisphaera</i>	1.09	0.030	0.005	0.007	0.61	degrader
Bacteroidetes	<i>Emticicia</i>	1.06	0.029	0.003	0.005	0.69	Fucosidase
Verrucomicrobia	<i>Opitutus</i>	1.06	0.029	0.012	0.019	0.66	H ₂ producer
Planctomycetes	<i>Rubinisphaera</i>	0.99	0.027	0.004	0.006	0.74	CO ₂ prod.
Elusimicrobia	<i>Endomicrobium</i>	0.91	0.024	0.005	0.008	0.63	VFA
Euryarchaeota	<i>Methanocaldococcus</i>	0.90	0.023	0.004	0.006	0.67	Methanogen (H)
Euryarchaeota	<i>Methanococcus</i>	0.86	0.023	0.011	0.015	0.74	Methanogen (H)
Euryarchaeota	<i>Methanosphaera</i>	0.86	0.023	0.032	0.041	0.77	Methanogen (H)

833 VIP: Variable importance for projection, Coef.: Coefficient, AA: Amino acids metabolim,
834 AOB: Ammonia-oxidizing bacteria, NOB: Nitrite-oxidizing bacteria, SOB: Sulphur-
835 oxidizing bacteria, ox.: Oxidizer, Methanogen (H): Hydrogenotrophic pathway, Methanogen
836 (M): Methylotrophic methanogenic pathway, VFA: Volatile Fatty Acids, N.I.: No
837 information.

838 Table 2: PLS results identifying the most important functional genes affecting methane
 839 emissions

KEGG ID	Function	VIP	Coef.	Mean Low CH4	Mean High CH4	L/H CH4 ratio
Increasing effects on methane emissions						
K06937	7,8-dihydro-6-hydroxymethylpterin dimethyltransferase	1.26	0.096	0.007	0.018	0.36
K00046	gluconate 5-dehydrogenase	1.13	0.089	0.067	0.098	0.68
K02117	V-type H ⁺ -transporting ATPase subunit A	1.07	0.067	0.135	0.216	0.62
K02118*	V-type H ⁺ -transporting ATPase subunit B	1.02	0.057	0.120	0.189	0.63
K00584* ^a	tetrahydromethanopterin S-methyltransferase subunit H	1.00	0.053	0.049	0.103	0.48
K00203 ^a	formylmethanofuran dehydrogenase subunit D	0.99	0.046	0.017	0.032	0.53
K00200* ^a	formylmethanofuran dehydrogenase subunit A	0.99	0.042	0.066	0.125	0.53
K00150	glyceraldehyde-3-phosphate dehydrogenase (NAD(P))	0.98	0.042	0.026	0.056	0.47
K01499 ^a	methenyltetrahydromethanopterin cyclohydrolase	0.97	0.037	0.040	0.079	0.50
K00169* ^a	pyruvate ferredoxin oxidoreductase, alpha subunit	0.97	0.034	0.032	0.062	0.52
K00580* ^a	tetrahydromethanopterin S-methyltransferase subunit D	0.95	0.031	0.021	0.045	0.47
K00400* ^a	methyl coenzyme M reductase system, component A2	0.95	0.027	0.022	0.047	0.48
K00170* ^a	pyruvate ferredoxin oxidoreductase, beta subunit	0.94	0.032	0.023	0.044	0.52
K13812* ^a	bifunctional enzyme Fae/Hps	0.94	0.029	0.031	0.062	0.50
K14128* ^a	F420-non-reducing hydrogenase subunit G	0.93	0.032	0.046	0.078	0.59
K02303	uroporphyrin-III C-methyltransferase	0.93	0.060	0.004	0.010	0.44
K14120*	energy-converting hydrogenase B subunit K	0.92	0.064	0.005	0.010	0.46
K00123* ^a	formate dehydrogenase, alpha subunit	0.92	0.007	0.126	0.206	0.61
K00201* ^a	formylmethanofuran dehydrogenase subunit B	0.91	0.028	0.091	0.155	0.58
K01959	pyruvate carboxylase subunit A	0.91	0.014	0.027	0.051	0.53
K00581* ^a	tetrahydromethanopterin S-methyltransferase subunit E	0.90	0.001	0.055	0.094	0.58
K00672 ^a	formylmethanofuran--tetrahydromethanopterin N-formyltransferase	0.89	0.003	0.024	0.056	0.43

K00399 * ^a	methyl-coenzyme M reductase alpha subunit	0.89	0.003	0.137	0.223	0.61
K01673	carbonic anhydrase	0.89	0.059	0.007	0.014	0.46
K00205 ^a	formylmethanofuran dehydrogenase subunit F	0.86	0.040	0.012	0.025	0.47
Low effects on methane emissions						
K03389	heterodisulfide reductase subunit B	1.13	-	0.047	0.069	0.68
K00440 *	coenzyme F420 hydrogenase alpha subunit	1.10	-	0.039	0.059	0.66
K00320	coenzyme F420-dependent N5,N10-methenyltetrahydromethanopterin reductase	1.03	-	0.076	0.109	0.70
K14123 * ^a	energy-converting hydrogenase B subunit N ¹	1.02	-	0.011	0.021	0.51
K00202 *	formylmethanofuran dehydrogenase subunit C ¹	1.01	-	0.043	0.066	0.65
K14101	energy-converting hydrogenase A subunit J ¹	1.00	-	0.007	0.012	0.55
K00125 * ^a	formate dehydrogenase, beta subunit ¹	1.00	-	0.051	0.082	0.62
K00401	methyl-coenzyme M reductase beta subunit ¹	1.00	-	0.089	0.135	0.66
K07388	hydrogenase expression/formation protein	0.95	-	0.019	0.031	0.60
K00577 * ^a	tetrahydromethanopterin S-methyltransferase subunit A ¹	0.94	-	0.029	0.057	0.51
K03390	heterodisulfide reductase subunit C	0.93	-	0.023	0.041	0.56
K00402	methyl-coenzyme M reductase gamma subunit ¹	0.86	-	0.051	0.076	0.67

840 ¹: Potential reasons for unexpected negative coefficients will be addressed in the discussion.

841 *: Genes also identified in the network analysis.

842 ^a: Genes previously identified in Roehle et al. (2016) as biomarkers of methane emissions.

843 VIP: Variable Importance in Projection. Coef.: Coefficient.

844

845

846

847

848

849

850

851 Table 3: PLS analysis comparing potential biomarkers correlated with CH₄ emissions.

Factor	VIP	Coefficient	Information
Phylogenetic factor¹			
<i>Methanotrorris</i>	1.69	0.14	Hydrogenotrophic methanogen
<i>Methanobrevibacter</i>	1.37	0.12	Hydrogenotrophic methanogen
<i>Methanocaldococcus</i>	1.25	0.11	Hydrogenotrophic methanogen
<i>Methanohalophilus</i>	1.09	-0.09	Methylotrophic methanogen
<i>Faecalitalea</i>	0.99	-0.09	AA
<i>Dorea</i>	0.90	-0.08	Acetogen
<i>Colwellia</i>	0.88	0.04	Alkane degrader
<i>Opitutus</i>	0.88	-0.03	H ₂ producer
<i>Singulisphaera</i>	0.88	-0.02	Degrader
<i>Isosphaera</i>	0.85	-0.03	Degrader
Microbial community factor			
PCoA-2 ²	1.55	-0.13	
Met Shannon Even ²	1.33	-0.12	Methanogen evenness
Met Shannon Div ²	1.32	-0.12	Methanogen diversity
A:B	0.88	-0.01	Archaea:Bacteria ratio
PCoA-1 ²	0.86	-0.05	
Metagenomics factor³			
K00672	1.32	-0.05	formylmethanofuran-tetrahydromethanopterin N-formyltransferase
K00581	1.08	-0.02	tetrahydromethanopterin S-methyltransferase subunit E
K00150	1.06	-0.02	glyceraldehyde-3-phosphate dehydrogenase (NAD(P))
K01959	1.02	-0.02	pyruvate carboxylase subunit A
K00580	1.00	-0.01	tetrahydromethanopterin S-methyltransferase subunit D
K01499	0.95	0.00	methenyltetrahydromethanopterin cyclohydrolase
K00584	0.94	0.00	tetrahydromethanopterin S-methyltransferase subunit H
K01673	0.93	-0.03	carbonic anhydrase
K13812	0.93	0.00	bifunctional enzyme Fae/Hps
K00123	0.92	0.01	formate dehydrogenase, alpha subunit
K00400	0.89	0.01	methyl coenzyme M reductase system, component A2
K00402	0.89	-0.07	methyl-coenzyme M reductase gamma subunit
K00399	0.89	0.00	methyl-coenzyme M reductase alpha subunit
K02118	0.87	0.01	V-type H ⁺ -transporting ATPase subunit B
K00200	0.87	0.02	formylmethanofuran dehydrogenase subunit A
K00201	0.86	0.01	formylmethanofuran dehydrogenase subunit B

K14128	0.85	0.03	F420-non-reducing hydrogenase subunit G
K00169	0.84	0.03	pyruvate ferredoxin oxidoreductase, alpha subunit
K00170	0.84	0.02	pyruvate ferredoxin oxidoreductase, beta subunit
K02117	0.84	0.02	V-type H ⁺ -transporting ATPase subunit A
K00203	0.82	0.04	formylmethanofuran dehydrogenase subunit D
K06937	0.80	0.03	7,8-dihydro-6-hydroxymethylpterin dimethyltransferase

852 ¹Value based on the relative abundance of the microbial genera identified as significantly
853 correlated by PLS, ²Data obtained by calculating the Shannon diversity indices or doing a
854 PCoA on the relative abundance of the microbial phyla, ³Value based on the relative
855 abundance of the genes identified as significantly correlated by PLS.

In review

Figure 1.JPEG

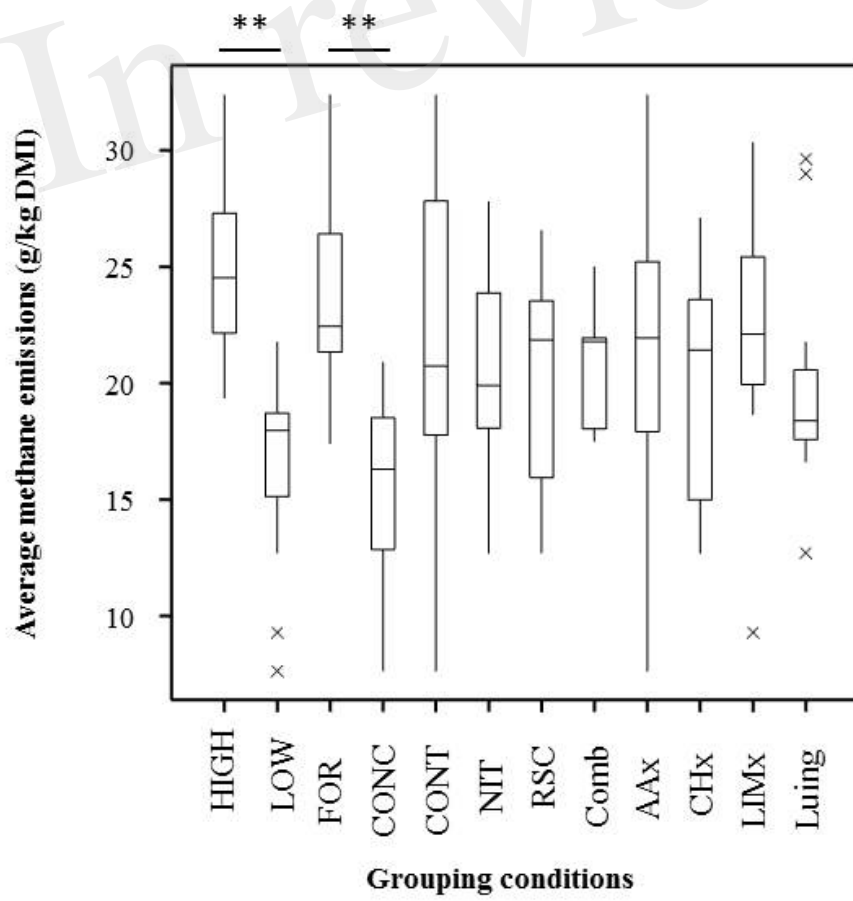


Figure 2.JPEG

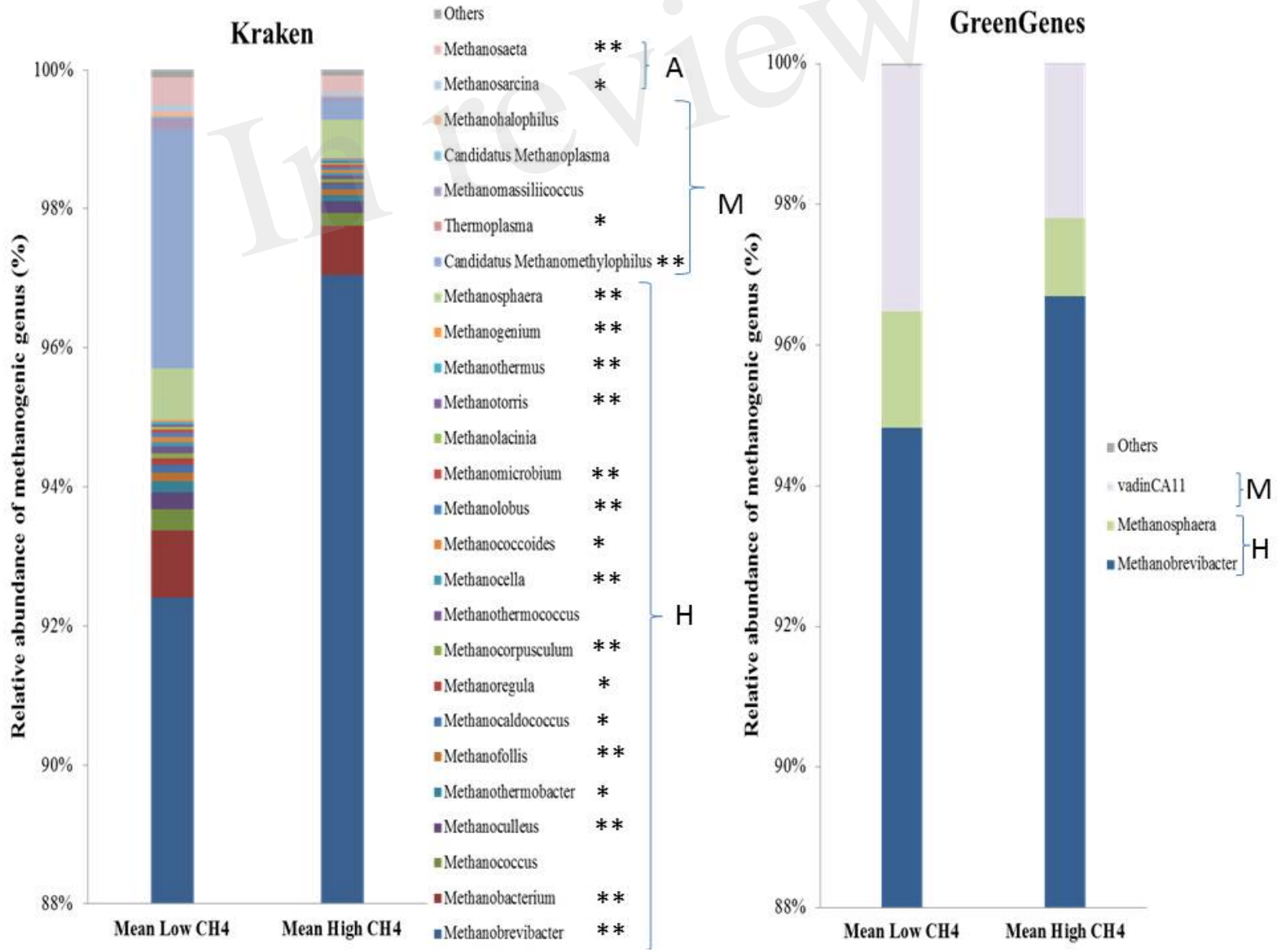


Figure 3.JPEG

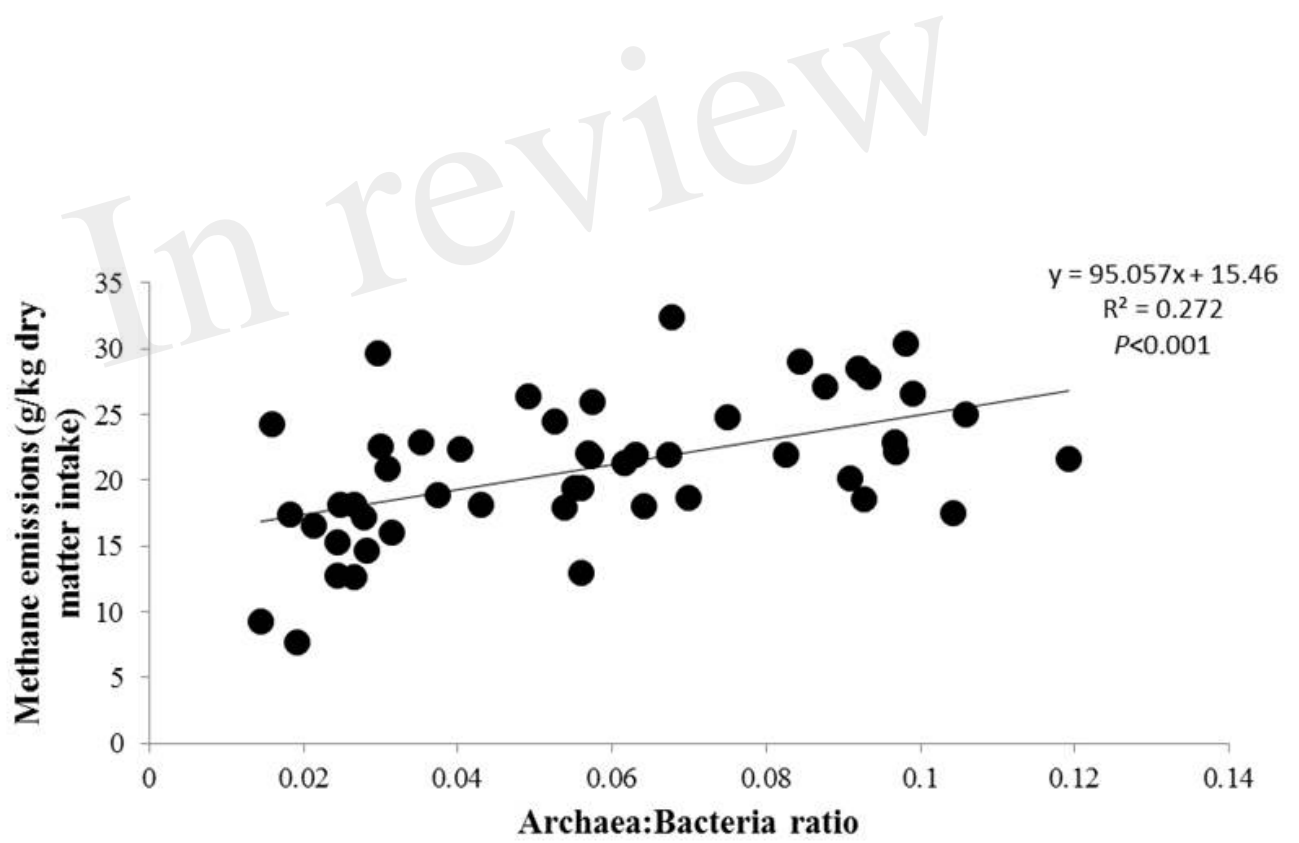
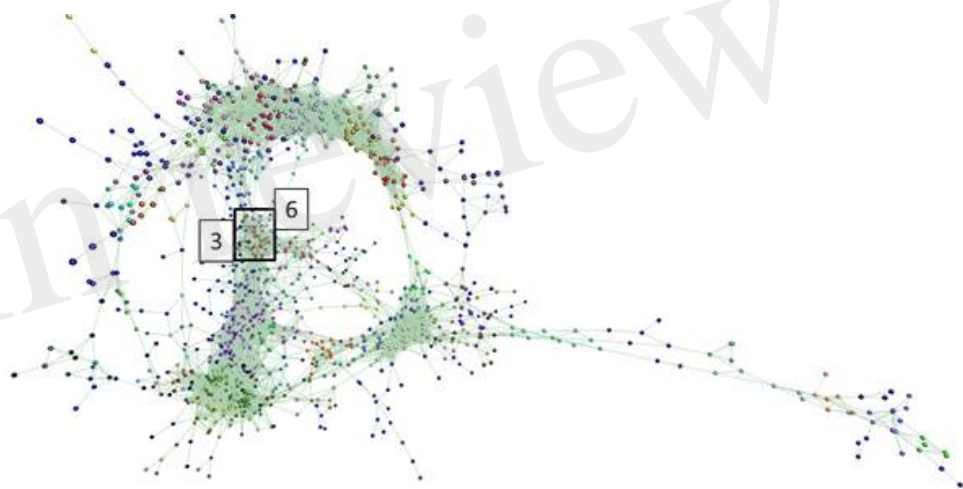
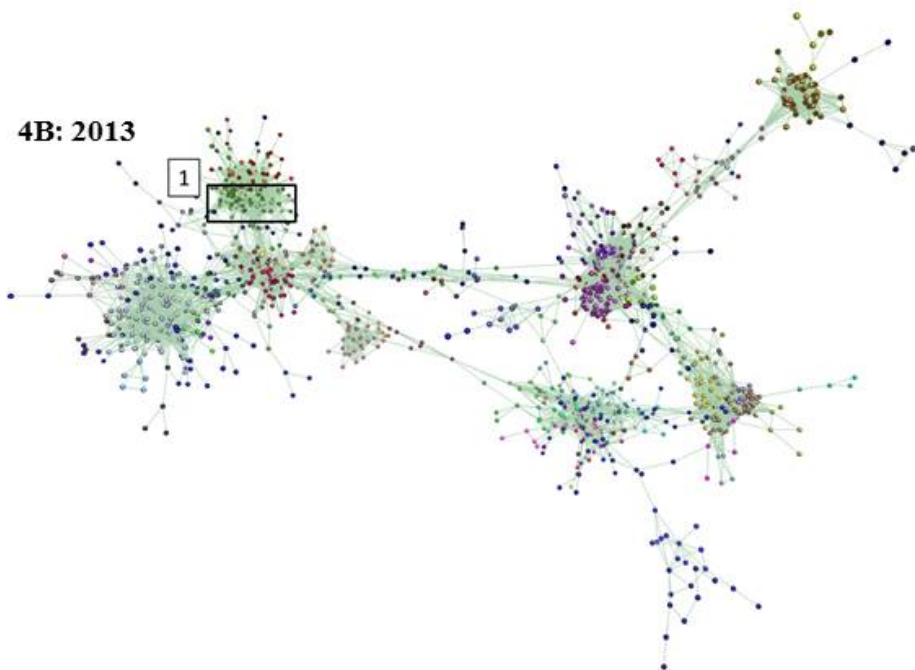


Figure 4.JPEG

4A: 2011



4B: 2013



4C: 2014

