Understanding the relationships between rumen microbiome genes and metabolites to be used for prediction of cattle phenotypes

Mengyuan Wang[1], Huiru Zheng[1]*, Haiying Wang[1], Richard J. Dewhurst[2], Rainer Roehe[2]
[1]School of Computing, Ulster University, United Kingdom
[2]Future Farming Systems, Scotland's Rural College, Edinburgh, United Kingdom
*Email: h.zheng@ulster.ac.uk

Abstract

The growing world population is facing increased future nutritional needs for meat and milk which need to be produced with minimal environmental impact, e.g. reduced methane emissions from ruminants. The combination of metagenomics and metabolomics can be effectively applied to understand rumen microbial gene expression, metabolic mechanisms that affect methane emissions and to address the challenges of ruminant production. Using 36 rumen samples derived from two omics studies, we conducted an in-depth analysis of the differences in diets and methane emissions from rumen metabolites and microbial genes. The top five integrals with significant (P<0.0001) differences in terms of their intensity measured across sample groups were found to be the same when samples were divided based on diet treatments and methane emissions. Based on the combination of statistical analysis and network approaches, this paper investigates the relationships between rumen microbial genes and integrals associated with metabolites which could be used for prediction of cattle phenotypes. Up to 98% of microbial genes and metabolites have no significant (P>0.05) linear correlation. The sample correlation network constructed using both integrals associated with metabolites and relative abundances of 20 microbial genes associated with methane emission exhibited a highly modular structure, which forms well-separated clusters according to different diet treatments. The evidence from this research confirmed the response of rumen microbes to different basal diets, and these activities subsequently affect methane emissions.

1 Introduction

Since 2006, atmospheric methane concentrations have increased at a rate of 0.4% per year [1]. Therefore, reduction of methane emissions in agriculture is of great interest for research. Rumen microbes interact closely to digest fibre structure, whilst providing metabolic energy to the host and producing methane under the action of archaea. This is a natural process responsible for one-third of methane from agriculture [2], [3]. Understanding multi-omics interactions is key to controlling rumen methane emission while meeting the growing demand for ruminant proteins.

Metagenomics is a genomic strategy to understand genetic composition and community function of all microorganisms contained in environmental samples [4]. However, the research found that the genome and proteome are still difficult to explain the interaction and activities involved in the metabolism pathway [5]. Metabolites are the final product of the entire cellular biological process [6], and the quantitative levels of metabolites could reflect the change of microbiological systems. Compared with genomics and proteomics, metabolomics can provide more information relating to the response of organisms to environmental stimulus and the metabolic pathways [7].

Bioinformatics techniques are widely used for analyzing large-scale multi-omics datasets [8]. The network-based approaches have shown great potential in integrating different omics information while being capable of capturing the fundamental properties of the microbial ecosystem including functional similarity and metabolic processes [9-10]. They have proven successful in bringing clarity to the complex relationship of microbes. For example, Wang et al. identified traits mircobial genes by constructed a co-abundance network and determined the threshold of the co-abundance network through a random forest algorithm, which improved the accuracy of the model [11-12].

Based on the combination of statistical analysis and network approaches for the study of 36 rumen samples derived from two omics studies, this paper investigated the relationships between rumen microbial genes and metabolites which could be used for prediction of cattle phenotypes (such as methane emission). The rest of the paper was organized as follows. Section II described the datasets and the methodology of this study. Experimental results were presented in Section III, followed by discussion and conclusions.

2 Data and Methodology

2.1 Dataset Description

The experiment conducted by the Beef and Sheep Research Centre of Scotland's Rural College (SRUC, Edinburgh, UK). This is a 2 × 2 factorial rotational experiment of genotypes and diets which was designed by Roehe et al. [9]. The cattles in the experiment were offered two complete diets which consisting (g/kg Dry Matter Intake) of 500 forage to 500 concentrate (FOR) or 80 forage to 920 concentrate (CONC).

Data used in this study included methane emissions, rumen microbial gene abundance, and integrals associated with rumen metabolite. 1) Methane emissions: During the experiment, methane emissions (g/kg Dry Matter Intake) from cattle were measured individually for 48h in respiration chambers as described in Rooke et al.[12]. Rumen fluid was extracted for next generation sequencing analysis and NMR analysis. The corresponding data of rumen microbial gene and rumen metabolite were provided by SRUC [13-16]. 2)Metagenomics data: the 1461 microbial gene with the relative abundance greater than 0.001 were selected. 3)Metabolomics data which were derived based on the peak of ^{1}H NMR (Nuclear Magnetic Resonance spectroscopy), which is in correspondence with the hydrogen atom of each signal in the sample. The relative intensity of the signals in the NMR analysis map reflects the relative content of each component in the sample. In this paper, 128 integrals obtained by NMR analysis were used to represent different metabolite concentrations. Among them, 10 integrals were identified for the corresponding metabolites. After removing the samples with missing metabolite data, a total of 36 samples were included in the study [12].

2.2 Data Analysis

The framework used in this study was illustrated in **Figure 1**. The study is divided into three parts. Firstly, the t-test was used to verify whether there are significant differences in methane emissions and metabolites when cattle were fed different diets. After that, the correlation between microbial genes and integrals were investigated. The significance of correlation coefficient was corrected by using Bonferroni method [17]. The network approach was applied to construct a sample correlation network based on integrals and relative abundance of microbial genes associated with methane emission, where nodes represent samples and the length of the edges indicates the strength of their correlation. The thresholds of correlation to construct the network were manually determined via correlation kurtosis distribution.

In this research, the t-test was implemented by using R 3.5.1 and the correlation between microbial genes and metabolites was calculated using the Expression Correlation plugins of Cytoscape [18]. The venny plot was created using the Venny2.1 platform [19]. The network visualization is through Cytoscape 3.7.1 [18].

3 Results

3.1 Factors Influencing the Abundance of Rumen Metabolites

Based on methane emission levels, the samples were divided into high and low groups when the difference between the average of the two groups reached the maximum. As expected, the two groups had significant (P<0.05) differences (**Figure 2**). When t test was carried out on samples of different diets, the average methane emission of FOR group was significantly (P<0.05) higher than the CONC group (**Figure 2**).

The t-test was carried out on all 128 integrals associated with metabolites. Interestingly, the top five integrals with significant differences (P<0.0001) in terms of their intensity measured across sample groups (diet treatments and methane emission groups) were found to be the same (**Table 1**); however, their association with metabolites are still unknown. Between FOR and CONC groups, there are 83 integrals exhibiting significant (P<0.05) difference, while only 62 found to be significantly (P<0.05) different in terms of methane emission. There were 58 integrals found to be significantly (P<0.05) different in both diets and methane emission groups (**Figure 3**). In addition, one integral associated with propionate was significantly (P<0.05) different between both diet treatments and methane emission groups. One butyrate signal and two propionate signals were significantly (P<0.05) different between two diet groups. In terms of the correlation between integrals and methane emission, only 10 integrals were found to be positive correlation with methane emission, and 92% of integrals were negatively correlated with methane emission.

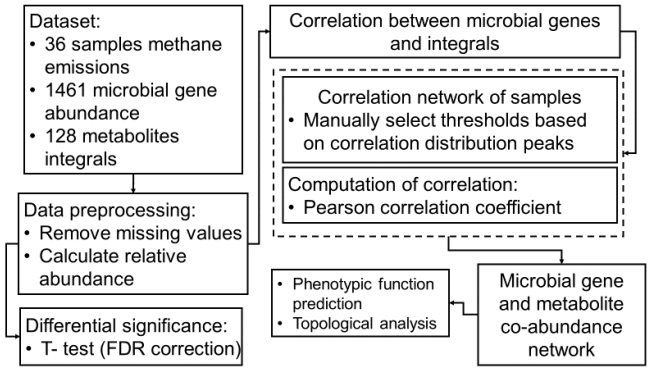

Figure 1 A workflow to illustrate the key steps of this research

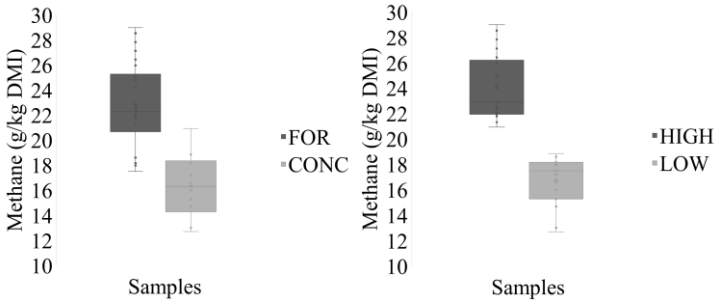

Figure 2 Factors Affecting Methane Emissions

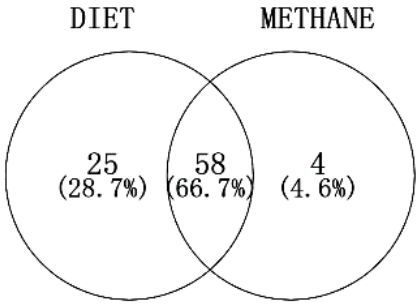

Figure 3 Venny Plot of Metabolites from Different Groups

Table 1 The Top 5 Significantly Different Integrals of Diets Group and Methane Emission Group

Diet treatment		Methane emission level	
Metabolites	*P* value	Metabolites	*P* value
Integral127	0.00000	Integral127	0.00000
Integral21	0.00000	Integral21	0.00000
Integral61	0.00000	Integral61	0.00000
Integral66	0.00000	Integral51	0.00000
Integral51	0.00000	Integral66	0.00000

3.2 Correlation between Rumen Microbial Genes and Integrals Associated with Metabolites

The linear correlation coefficients between 1461 microbial genes and 128 integrals were all lower than 0.5 in which 98% exhibited no significant (P>0.05) difference. About 80% correlation were between -0.3 and 0.3. The top pairs in terms of absolute correlation between metabolites and microbial genes were listed in **Table 2**, in which only correlation between K03415 and Integral59 was found to be significant (P<0.05).

Table 2 Correlation coefficient between microbial genes and integrals of the top 5

Microbial Gene	Metabolites	Correlation coefficient	*P* Value
K03415	Integral59	0.5019261	0.006
K01493	Integral73	0.5007088	0.09
K01813	Integral85	0.5005219	0.091
K01783	Integral11	0.4997033	0.669
K06987	Integral18	0.4975293	0.894

3.3 Correlation network analysis

Sample correlation networks were constructed based on the integrals and abundance of microbial methane emission functional genes. As shown in **Figures 4** (A) and (B), the correlation networks derived from the abundance of microbial methane emission functional genes and integrals alone failed to differentiate between samples with different diet treatments. However, when combining the abundance of microbial methane emission functional genes and integrals associated with metabolites, samples with FOR are clearly separated from the rest of samples **Figure 4** (C). Interestingly the FOR sample (RR0003) is grouped with 8 CONC samples which deserves further investigation. Samples with similar methane emission levels does not show a closer correlation.

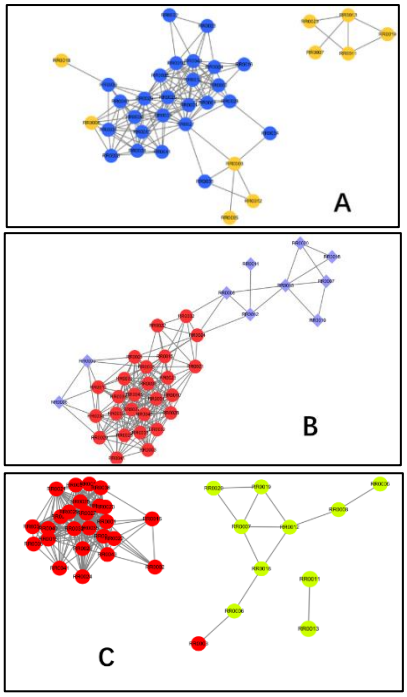

Figure 4 Different colors for different diets: (A) Sample correlation network based on methane emission functional microbial genes; nodes:36, edges:369, thesrhold:0.91, blue nodes: FOR, yellow nodes: CONC. (B) Sample correlation network based on all metabolites; nodes:36, edges:210, thesrhold:0.995, red nodes: FOR, purple nodes: CONC. (C) Sample correlation network based on the combination of metabolites and methane emission functional microbial genes; nodes:36, edges:225, thesrhold:0.98, red nodes: FOR, green nodes: CONC

4 Conclusions and Discussion

It has been shown that integration of different omics information has potential to provide important insights into response mechanisms of rumen microbes ranging from different genotype cattle to the environmental stimuli such as different diet formulation [1]. In order to explore how effectively combine the information of microbial genes and metabolite for future bovine phenotypic prediction, this study studies the relationships between integrals which associated with metabolites and microbial genes using 36 bovine samples with two diet treatments and different methane emission levels.

1) As expected, we confirmed that the different basal diets strongly and significantly influenced the methane emissions, and the methane emission level were significantly

(P<0.05) higher in the FOR group than in the CONC group. Significant (P<0.05) differences of metabolites between the high and low methane emission groups were observed. Similarly, significant (P<0.05) differences between metabolites were also found in different diet groups. It is most likely that the basal diet changed the environmental conditions in the rumen (such as pH) to affect microbial activity, which ultimately affected the methane emissions. When analysed the significant difference of the metabolite integrals respectively in the methane emission groups and the diets groups, it was found the top five of each group are the same. And only 8% of all the integrals had positive correlations with methane emissions. This provides clues to find key metabolites for the prediction of rumen methane emissions.

2) There were weak correlations between metabolites and microbial genes. The absence of strong correlations may be due to the absorption, degradation or utilization of metabolites by a series of metabolisms in the rumen. The current study is based on the analysis of linear correlation relationship between metabolites and microbial genes. In the future, the correlation based on nonlinear parameters such as distance or mutual information will be investigated. Using different nonlinear calculation methods (e.g. mutual information) or methods designed for microbiome correlation analysis (e.g. CCLasso or SparCC) [20-21] is another direction for future research.

3) Most FOR samples were clearly separated from the CONC samples when the correlations were calculated based on both metabolites and microbial methane emission functional genes. It implies that the combination of metabolites and microbial methane emission functional genes magnified the differences caused by diets. Further in-depth investigation is required to improve the rule.

Acknowledgement

This research is jointly supported by Ulster University and Scotland's Rural College, U.K. and partially supported by the MetaPlat project, (www.metaplat.eu), funded by H2020MSCA-RISE-2015.

5 Literature

[1] S.A. Huws, C.J. Creevey, L.B. Oyama, I. Mizrahi, S.E. Denman, M. Popova, R. Muñoz-Tamayo, E. Forano, S.M. Waters, M. Hess, I. Tapio, H. Smidt, S.J. Krizsan, D.R. Yáñez-Ruiz, A. Belanche, L. Guan, R.J. Gruninger, T.A. McAllister, C.J. Newbold, R. Roehe, R.J. Dewhurst, T.J. Snelling, M. Watson, G. Suen, E.H. Hart, A.H. Kingston-Smith, N.D. Scollan, R.M. do Prado, E.J. Pilau, H.C. Mantovani, G.T. Attwood, J.E. Edwards, N.R. McEwan, S. Morrisson, O.L. Mayorga, C. Elliott, and D.P. Morgavi: Addressing Global Ruminant Agricultural Challenges Through Understanding the Rumen Microbiome: Past, Present, and Future. Front Microbiol. vol. 9, 2018.

[2] W. Lan and C. Yang: Ruminal methane production: Associated microorganisms and the potential of applying hydrogen-utilizing bacteria for mitigation. Science of The Total Environment. vol. 654, pp. 1270–1283, 2019.

[3] C.-A. Duthie, S.M. Troy, J.J. Hyslop, D.W. Ross, R. Roehe, and J.A. Rooke: The effect of dietary addition of nitrate or increase in lipid concentrations, alone or in combination, on performance and methane emissions of beef cattle. animal. vol. 12, no. 02, pp. 280–287, 2018.

[4] F. Browne, H. Wang, H. Zheng, R. Roehe, R.J. Dewhurst, and P. Walsh: A network analysis of methane and feed conversion genes in the rumen microbial community. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1477–1484 (2016).

[5] S.E. Denman, D.P. Morgavi, and C.S. McSweeney: Review: The application of omics to rumen microbiota function. Animal. vol. 12, no. s2, pp. s233–s245, 2018.

[6] R. Wohlgemuth: Horizons of Systems Biocatalysis and Renaissance of Metabolite Synthesis. Biotechnology Journal. vol. 13, no. 6, pp. 1700620, 2018.

[7] T. Hao, D. Wu, L. Zhao, Q. Wang, E. Wang, J. Sun: The Genome-Scale Integrated Networks in Microorganisms. Front Microbiol, Vol. 9,2018.

[8] D.K. Barupal, S. Fan, and O. Fiehn: Integrating bioinformatics approaches for a comprehensive interpretation of metabolomics datasets. Current Opinion in Biotechnology. vol. 54, pp. 1–9, 2018.

[9] R. Roehe, R.J. Dewhurst, C.-A. Duthie, J.A. Rooke, N. McKain, D.W. Ross, J.J. Hyslop, A. Waterhouse, T.C. Freeman, M. Watson, and R.J. Wallace: "Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance." PLoS Genet. vol. 12, no. 2, pp. e1005846, 2016.

[10] R.J. Wallace, J.A. Rooke, C.-A. Duthie, J.J. Hyslop, D.W. Ross, N. McKain, S.M. de Souza, T.J. Snelling, A. Waterhouse, and R. Roehe: Archaeal abundance in post-mortem ruminal digesta may help predict methane emissions from beef cattle. Scientific Reports. vol. 4, no. 1, 2015.

[11] H. Wang, H. Zheng, F. Browne, R. Roehe, R.J. Dewhurst, F. Engel, M. Hemmje, X. Lu, and P. Walsh: Integrated metagenomic analysis of the rumen microbiome of cattle reveals key biological mechanisms associated with methane traits. Methods. vol. 124, pp. 108–119, 2017.

[12] M.D. Auffret, R. Stewart, R.J. Dewhurst, C.-A. Duthie, J.A. Rooke, R.J. Wallace, T.C. Freeman, T.J. Snelling, M. Watson, and R. Roehe: Identification, Comparison, and Validation of Robust Rumen Microbial Biomarkers for Methane Emissions Using Diverse Bos-Taurus Breeds and Basal Diets. Front. Microbiol. vol. 8, 2018.

[13] S.M. Troy, C.-A. Duthie, J.J. Hyslop, R. Roehe, D.W. Ross, R.J. Wallace, A. Waterhouse, and J.A. Rooke: Effectiveness of nitrate addition and increased oil content as methane mitigation strategies for beef cattle fed two contrasting basal diets. J. Anim. Sci. vol. 93, no. 4, pp. 1815–1823, 2015.

[14] J.A. Rooke, R.J. Wallace, C.-A. Duthie, N. McKain, S.M. de Souza, J.J. Hyslop, D.W. Ross, T. Waterhouse, and R. Roehe: Hydrogen and methane emissions from beef cattle and their rumen microbial community vary with diet, time after feeding and genotype. Br. J. Nutr. vol. 112, no. 3, pp. 398–407, 2014.

[15] C.-A. Duthie, Rooke J.A., Troy S., Hyslop J.J., Ross D.W., Waterhouse A., and Roehe R.: Impact of adding nitrate or increasing the lipid content of two contrasting diets on blood methaemoglobin and performance of two breeds of finishing beef steers. Animal . vol. 10, no. 5, pp. 786–795, 2016.

[16] C.-A. Duthie, M. Haskell, J.J. Hyslop, A. Waterhouse, R.J. Wallace, R. Roehe, and J.A. Rooke: The impact of divergent breed types and diets on methane emissions, rumen characteristics and performance of finishing beef cattle. Animal. vol. 11, no. 10, pp. 1762–1771, 2017.

[17] J.M. Bland and D.G. Altman: Statistics notes: Multiple significance tests: the Bonferroni method. BMJ. vol. 310, no. 6973, pp. 170–170, 1995.

[18] A. Niissalo: Cytoscape and its plugins. 2007.

[19] J.C. Oliveros: Venny. An interactive tool for comparing lists with Venn's diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html, 2007-2015

[20] H. Fang, C. Huang, H. Zhao, and M. Deng: CCLasso: correlation inference for compositional data through Lasso. Bioinformatics. vol. 31, no. 19, pp. 3172–3180, 2015.

[21] J. Friedman and E.J. Alm: Inferring Correlation Networks from Genomic Survey Data. PLoSComputational Biology. vol. 8, no. 9, pp. e1002687, 2012.