VISUALISATION OF RUMEN MICROBIOME DATA

G.R. Wood[1], S. Kittelmann[2], C. Pinares-Patiño[2], K.G.Dodds[1], M.R. Kirk[2], S. Ganesh[2], S.M. Hickey[3], P.H. Janssen[2], J.C. McEwan[1] and S.J. Rowe[1]

[1]AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand
[2]AgResearch, Grasslands Research Centre, Palmerston North, New Zealand
[3]AgResearch, Ruakura Research Centre, Hamilton, New Zealand

SUMMARY

The rumen microbiome plays a key role in the production of methane, a critical greenhouse gas. Two graphical methods aiding visualization of rumen microbiome data are presented. Marker gene sequence data generated from 520 rumen samples from 260 sheep were used to estimate the relative abundance of bacteria. These were assigned to 54 bacterial taxa, which were then clustered into four groups of co-occurring taxa. Sheep were measured for methane emissions in open circuit respiration chambers. Heatmaps and simplex plots were used to visualise the rumen data and their relationship to methane emissions from the sheep. Although the relationship of these dimensions to methane emission was not clear-cut, the simplex graphic suggests that there is a continuum of low methane emitters across rumen profiles. The analysis indicated that the bacterial microbiome data set is broadly two-dimensional; two rumenotypes dominate. This in turn creates a challenge - to uncover the origins of this relative simplicity of rumen biology.

INTRODUCTION

Methane emission from ruminants (primarily sheep and cattle) contributes approximately 30% of annual anthropogenic greenhouse gas (GHG) production in New Zealand. Methane is produced in a two-stage process in the rumen: feed digestion aided by rumen bacteria yields hydrogen as a by-product, which in turn is converted to methane by methanogenic archaea.

The bacterial composition of the rumen microbiome can be determined through sequencing part of the gene coding for a slowly evolving ribosomal RNA. Sequenced sections are aligned to reference sequences of the gene in different bacteria and then assigned to bacterial taxa. This approach resulted in 54 taxa in the data set used in this study. The relative abundance of these taxa in the rumen of each animal can then be estimated. Taxa, once established, can be clustered if their relative abundances vary in concert across animals; in turn, this allows clustering of taxa. These co-occurring taxa are not necessarily related evolutionarily – instead, they are probably linked ecologically. A final research aim, of which the current research is a part, is to determine an animal measure which can be used to select for lower methane emission. It is known that animal genetics influences methane production (Pinares-Patiño *et al*. 2013) and that the rumen microbiome influences methane production (Kittelmann *et al*. 2014; Ross *et al*. 2013; Wallace *et al*. 2014).

The microbiome consists of multiple microbial species, each representing a dimension in the data set. Microbiome compositions within large animal groups are unlikely to form discrete clusters (where each point is a rumen sample), and so methods for simplifying these large data sets are useful in order to make progress. One route is to reduce the dimensionality – from 54 to a workable number. Classical methods are available, such as correspondence analysis. The aim here, however, is to present two simple visual methods, first a marriage of cluster analysis and heatmaps and second (related to correspondence analysis, (Greenacre 1983)) a marriage of cluster analysis and simplex plots. Each presents graphically, but in different ways, the relationship between animals, microbiome and methane production.

It is stressed that this paper focuses on exploratory graphical methods, designed to allow the

researcher to "see", and hence conjecture, relationships; confirmatory standard statistical tests are not conducted. Graphical methods are readily accessible; two examples are presented here with a view to providing useful tools to aid the progress of microbiome research.

MATERIALS AND METHODS

A full description of the selection lines used together with the materials and methods for measuring methane emissions can be found in Pinares-Patiño *et al*. (2011). Selection began in 2009 and the sample of sheep used in this study were from the 2011 and 2012 birth cohorts. Two rumen contents samples were collected by stomach tube from each of 260 New Zealand crossbred sheep at ~10 months of age. Rumen samples were collected 14 days apart approximately 18 hours after the last feed. The sheep were from selection lines phenotypically divergent for methane emissions per unit of dry matter intake (CH_4/DMI; g/kg) and were measured in open circuit respiration chambers for methane emissions continuously for 48 hours prior to each rumen sampling Pinares-Patiño *et al.* (2013). Sheep were offered a ration of lucerne pellets based on 2.0× maintenance energy requirements. Feed was offered twice daily and individual dry-matter intakes were recorded. Breeding values were estimated using ASReml 3 software (Gilmour 2009).

The method for assigning amplified 16S rRNA gene sequences to taxa is described by Kittelmann *et al*. (2014). The statistical package R was used for clustering and heatmaps and Matlab for the production of rotatable three-dimensional simplex plots (two representative screenshots are presented here). Simplexes can be drawn in each dimension; one-dimensional (line), two-dimensional (triangle); three-dimensonal (tetrahedron), etc. They provide the regions in which to picture "compositional" (summing to one) data, such as rumen relative abundances.

For both the heatmap and simplex plot, the bacterial taxa were reduced to four "rumenotypes" or clusters using *k*-means clustering; distance between taxa were established using the relative abundance profiles across animals. Four taxa clusters were chosen for two reasons: first they capture the bulk of the variation across taxa and second, this choice is the largest that can be readily pictured in simplex plot. It is important to stress that the rumenotypes cluster by co-occurrence across animals, and not by phylogenetic relationship of the bacterial taxa. Relative abundances of taxa in a cluster were summed and assigned to the four rumenotypes. For the heatmap, rows in the resulting 520×4 matrix of animals by rumenotypes were again clustered using *k*-means. Cluster centre values were log transformed to accentuate the differences in the animal groups seen in the heatmap.

RESULTS AND DISCUSSION

The two graphical approaches are now described, both employing clustering, with the first using heatmaps and the second using simplex plots. Each has advantages; the simplicity of the heatmap reveals the animal clusters readily whereas the greater detail of the simplex plot (which does not cluster animals) reveals the relationship between animal groups and methane emission more clearly.

Heatmaps. The resulting dendrogram of rumenotype clusters is shown against the columns, and the dendrogram of animal similarities by rumen bacterial community is shown on the left, in Figure 1. Finally, zero-centred breeding values for CH_4/DMI were added in the right column of Figure 1. Three aspects of this heatmap are noteworthy. First, the row profiles (animals) are broadly of two types, indicated by the row dendrogram. The row microbiome profiles run either "red, red, yellow, yellow" or "red, yellow, red, yellow". Second, the methane breeding values run counter to this division. The two lowest values, −0.037 and −0.016, lie in the lower and upper microbiome profile types. This suggests two broad bacterial community types, each with high and low methane variants. This will be seen again, in different form, in the second graphical presentation. Third, the relative abundance of Rumenotype4 is generally high.

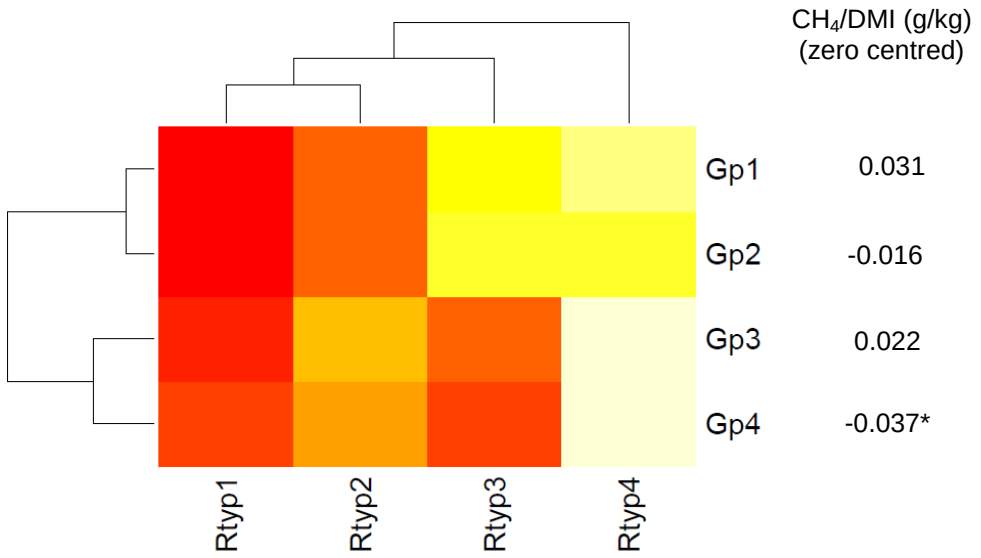

Figure 1. A heatmap (red low, yellow high) of relative abundance of bacterial clusters within the sheep microbiome. Four animal clusters (Gp1-Gp4, with animal counts of 70, 95, 214 and 141 respectively) by four microbiome clusters (Rumenotype1-Rumenotype4) are shown. Zero-centred breeding values for CH_4/DMI for each animal group (with a more negative breeding value corresponding to lower expected CH_4/DMI in progeny) are shown in the right hand column. The main conclusion is that high abundance of Rumenotype4 and moderately low abundance of Rumenotype3 can be associated with lower methane emission (the asterisked Gp4 animal row).

Simplex plots. Using the same co-occurring taxon clusters (Rumenotype1-Rumenotype4) but without animal clustering, we can plot all 520 sample relative abundance profiles in a three-dimensional simplex (a tetrahedron), colouring the resulting points according to level of methane emission. The resulting plot can be rotated; a selection of views is given in Figure 2. It suffices to plot the first three components of each point; this corresponds to mapping the tetrahedron spanned by the unit vectors in R^4 onto the unit vectors in R^3 together with the origin, in the canonical way.

Three main conclusions can be drawn from Figure 2. First, the dimensionality of the clustered microbiome is essentially two (seen in this approach in the planarity of the points). Second, the high relative abundance of Rumenotype4 (the (0,0,1) corner where many points lie). Third, lower CH_4/DMI (coloured red) occur in a continuum across the microbiome space, with some concentration at high levels of Rumenotype4.

Conclusions. We conclude with some overall remarks:

i) The microbiomes are essentially two-dimensional (this emerges from both graphical approaches). This is justified generally in that the four rumenotype clusters used capture the bulk of across animal variation.

ii) The relationship of these dimensions to methane emission is not yet clear cut using this data set; each graphic indicates that low emission occurs across the spectrum of animal groups.

iii) Each graphical approach offers advantages. For example, the heatmap shows the relative abundances in the animal groups more clearly while the simplex plot shows the two-dimensionality of the microbiome space more clearly.

iv) The challenge remaining is to provide a fundamental biological explanation for the core two-dimensionality of the rumen microbiome suggested here.

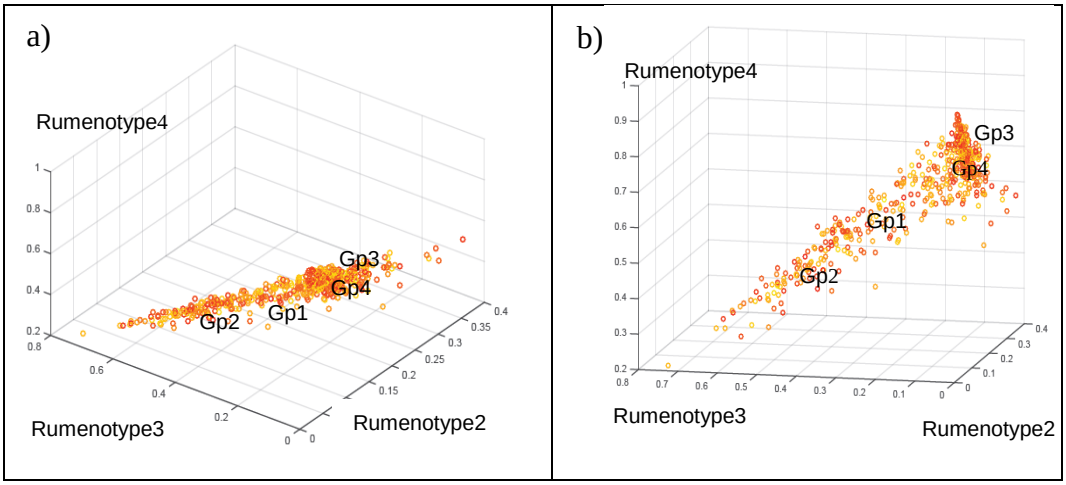

Figure 2. Tetrahedron plot of animal microbiome profiles. Each point represents the four-component (summing to one) relative abundances of Rumenotype1-Rumenotype4 for an animal microbiome sample. These are plotted within a three-dimensional tetrahedron, spanned by (0,0,0), (1,0,0), (0,1,0) and (0,0,1) ; colours represent level of CH_4/DMI (with red low and yellow high, but here note that these refer to level of methane emission, not relative abundance as in Figure 1). Four vertices of the unit cube represent extreme microbiomes (specifically, (0,0,0), (1,0,0), (0,1,0) and (0,0,1) correspond to all weight on Rumenotype1, Rumenotype2, Rumenotype3 and Rumenotype4 respectively) . Two perspectives are shown (the source object can be rotated in Matlab): in a) the planar nature is apparent while in b), the Rumenotype4 animals, with red (low) colour, show a microbiome region of low methane emission.

ACKNOWLEDGEMENTS

This study was funded by the New Zealand Pastoral Greenhouse Gas Research Consortium (PGgRc) and the New Zealand Agricultural Greenhouse Gas Research Centre (NZAGRC). Beef and Lamb New Zealand, AgResearch Limited, Lincoln University and On Farm Research are acknowledged for generating the animals and providing access to the Central Progeny Test flocks.

REFERENCES

Gilmour, A.R., Gogel, B.J., Cullis, B.R. and Thompson, R. (2009) ASReml User Guide Release 3.0. VSN International Ltd.

Greenacre, M. *Theory and Applications of Correspondence Analysis*. London: Academic Press (1983).

Kittelmann S., Seedorf H., Walters W.A. *et al.* (2013) *PLoS One* **8**, e47879.

Kittelmann S., Pinares-Patiño C.S., Seedorf H. *et al*. (2014) *PLoS One* **9**, e103171.

Pinares-Patiño C.S., McEwan J.C., Dodds K.G. *et al*. (2011) *Animal Feed Science and Technology* **166-167**:210-218.

Pinares-Patiño C.S., Hickey S.M., Young E.A. *et al*. (2013) *Animal* **7**: suppl 2: 316.

McDonald D., Price M.N., Goodrich J. *et al*. (2012) *The ISME Journal* **6**: 610.

Ross, E.M., Moate, P.J., Marett, L.C. *et al.* (2013) *PLoS One*, **8**, e73056.

Wallace R.J., Rooke J.A., Duthie C.A. *et al.* (2014) *Scientific Reports* **4**, 5892.