

Review: The application of omics to rumen microbiota function

S. E. Denman^{1†}, D. P. Morgavi² and C. S. McSweeney¹

¹CSIRO, Agriculture and Food, St. Lucia, QLD 4067, Australia; ²Institute National de la Recherche Agronomique, UMR1213 Herbivores, Clermont Université, VetAgroSup, UMR Herbivores, Clermont-Ferrand, France

(Received 19 June 2018; Accepted 27 August 2018; First published online 28 September 2018)

Rumen microbiome profiling uses 16S rRNA (18S rRNA, internal transcribed spacer) gene sequencing, a method that usually sequences a small portion of a single gene and is often biased and varies between different laboratories. Functional information can be inferred from this data, but only for those that are closely related to known annotated species, and even then may not truly reflect the function performed within the environment being studied. Genome sequencing of isolates and metagenome-assembled genomes has now reached a stage where representation of the majority of rumen bacterial genera are covered, but this still only represents a portion of rumen microbial species. The creation of a microbial genome (bins) database with associated functional annotations will provide a consistent reference to allow mapping of RNA-Seq reads for functional gene analysis from within the rumen microbiome. The integration of multiple omic analytics is linking functional gene activity, metabolic pathways and rumen metabolites with the responsible microbiota, supporting our biological understanding of the rumen system. The application of these techniques has advanced our understanding of the major microbial populations and functional pathways that are used in relation to lower methane emissions, higher feed efficiencies and responses to different feeding regimes. Continued and more precise use of these tools will lead to a detailed and comprehensive understanding of compositional and functional capacity and design of techniques for the directed intervention and manipulation of the rumen microbiota towards a desired state.

Keywords: rumen, microbial, metagenomics, metatranscriptomics, taxonomy

Implications

Demand for global animal protein and pressures for cropping land will combine to drive increased feed efficiency in ruminants from low quality cellulosic feeds. However, livestock production is responsible for a large proportion of global agricultural greenhouse gas emissions. Observing and understanding the highly complex relationships that influence the rumen microbiota are essential for designing methods to successfully intervene and manipulate the system towards a desired phenotype.

Introduction

The complex relationship between the host ruminant and its inhabitant microbiota has been the focus of research for decades, initially centred around identification of the types of microbiota that reside within the rumen through to a greater understanding of their functional contribution to the host's energy requirements. As new technology has become available, what originally involved the isolation and detailed studies of single strains in the laboratory has now moved to

large scale sequencing of 'total' rumen microbiota nucleic acids (metagenomics and metatranscriptomics), proteomics and metabolomics. Notwithstanding the limitations of these new techniques, which will be discussed in more detail throughout this review, the adoption of these techniques has been rapid and applied to most ruminant production systems. Initially to define the variance in rumen bacterial populations with diet shifts, which can lead to digestive disorders such as acidosis. Then in the last decade, the emphasis has been focussed around the understanding of the rumen microbiota's contribution to agricultural greenhouse gas emissions, predominantly methane. While now there is increased interest on defining the rumen microbiota of an efficient production animal (meat and dairy) and the influence of the host genetics on shaping the microbiota of the rumen (Jami *et al.*, 2014).

Characterising the rumen microbiota using taxonomic marker genes

Considerations and limitations

Metataxonomics has been recommended as the term for defining high throughput sequencing analysis of amplified

[†] E-mail: stuart.denman@csiro.au

taxonomic marker genes, whereas metagenomics refers to the use of shotgun sequencing approaches to characterise the potential function of the microbiota based on their genomes (Marchesi and Ravel, 2015). Identifying the microbiota that resides within the rumen and variations to the composition in response to perturbations is critical for the development of our understanding of the complex dynamics that exist within the rumen environment. Metataxonomics is routinely employed in ruminant studies and is relatively inexpensive as it easily allows for the pooling of many samples and reasonable sequence depth per analysis compared with other metaomic techniques. General primer sets are used to amplify sequences representing the archaeal and bacterial members by targeting the 16S rRNA gene, whereas the 18S rRNA gene is used for eukaryotic targets, essentially protists. Due to the high degree of sequence similarity found within the fungal 18S rRNA, the anaerobic fungal populations require targeting of the length polymorphic internal transcribed spacer sequence that lies between the ribosomal genes (Dore and Stahl, 1991; Kittelmann *et al.*, 2012). Although more recently, focus has shifted towards the D1/D2 region at the 5' end of the 28S rRNA gene (Edwards *et al.*, 2017). The choice of DNA extraction method, amplification primer sets, sequencing platform and bioinformatics workflows will all influence the final analysis (Gantner *et al.*, 2011; Henderson *et al.*, 2013; Klindworth *et al.*, 2013). The variance in these methods and how they are applied makes reconciling results between research groups difficult, although using consistent methods and control samples can produce valid observations (Henderson *et al.*, 2013).

Phylogenetic gene analysis is based on the sequence similarity or, more correctly, evolutionary divergence of sequences between defined taxonomic units. Initially, and at least for the full length 16S rRNA gene, a value of 97% sequence similarity was used to define a species level rank (Stackebrandt and Goebel, 1994). However, this has now been redefined to a recommended value of 98.5% (Konstantinidis and Tiedje, 2007). This does not hold true for the shorter sequences generated from the next generation sequencing technologies and values of 99% to 100% are more common for defining an operational taxonomic unit (OTU) rather than to a 97% species rank (Callahan *et al.*, 2016). Using a value of 100% is also likely to generate multiple OTUs from the same species due to species which possess multiple copies of the 16S rRNA gene and that these copies are known to not be identical (Větrovský and Baldrian, 2013). The choice of variable region targeted along with relaxing the identity threshold to 99% to allow for possible polymorphism effects may mitigate these. The most popular variable region currently being targeted is the V4 region, with primers that cover both bacterial and archaeal populations, producing an amplicon size amenable to the Illumina MiSeq sequencing platform (Kozich *et al.*, 2013).

As OTUs are defined by sequence similarity, it is imperative that errors arising from sample processing and sequencing platforms are minimised and removed (Kunin *et al.*, 2009). Software methods to correct for sequencing platform errors

are available and widely employed to improve the quality of sequence data. (Kunin *et al.*, 2009; Quince *et al.*, 2009; Bragg *et al.*, 2012, Callahan *et al.*, 2016). The inclusion of appropriate negative controls and standards across multiple experiments should be included to account for contaminants and variance in sample processing.

The most widely used analysis for rumen environments involves the 16S rRNA gene, for which curated databases such as Greengenes, RDP and SILVA exist (Cole *et al.*, 2003; DeSantis *et al.*, 2006; Pruesse *et al.*, 2007). The databases are constantly improving by including measures that account and remove inaccuracies in the data sets, especially around chimeric sequences and taxonomic nomenclature. (Yilmaz *et al.*, 2014; Balvočiūtė and Huson, 2017; Edgar, 2018). However, a large proportion of the databases are derived from environmental sequences and are generally not full length. Taxonomy for these sequences is therefore predicted using various methods such as Bayesian or sequence alignments to curated trees (Wang *et al.*, 2007; McDonald *et al.*, 2012). Many of the characterised isolates are the sole representative for an identified genus, thus making genus the lowest level of rank available for many classifiers.

Due to the intense interest in methanogenic archaeal populations in relation to their contribution to greenhouse gas emissions from livestock, a highly resolved taxonomic database focussed on gut isolates has been generated (RIM-DB) (Seedorf *et al.*, 2014).

Taxonomic inconsistencies and variance in accuracy of rank classification between reference sets poses a challenge for comparing results found within the literature. Although a common agreed workflow would minimise these issues, it is unlikely that this will eventuate. Agreed minimum descriptions of sample collection and data analysis have been requested MIMARKS (Yilmaz *et al.*, 2011).

The nature of these sequencing methods produces compositional data that restrict analysis to relative abundance methods and excludes standard Pearson and Spearman correlation analysis (Pearson, 1897; Lovell *et al.*, 2015; Gloor *et al.*, 2017). Alternative methods such as Aitchison and PhILR for Beta diversity analysis and ϕ to describe the strength of proportionality between two variables for describing correlations are available (Aitchison *et al.*, 2000; Lovell *et al.*, 2015; Silverman *et al.*, 2017). Furthermore, issues around variance in 16S rRNA gene copy numbers between species have not been adequately addressed and cannot be accurately determined for OTUs that are not represented by characterised species (Louca *et al.*, 2018). Further confounding the analysis is sequencing depth between samples, abundance issues arise around the variation in the number of sequences obtained from a given sample. Rarefying the data to a defined level across all samples preferentially excludes lower abundant OTUs leading to a loss of precision, whereas those that use the entire data set must account for the magnitude of sequence depth between samples and usually employ a transformation or scaling method (Anders and Huber, 2010; Robinson *et al.*, 2010; McMurdie and Holmes, 2014; Weiss *et al.*, 2017).

Methods for the identification of OTUs that are significantly associated with a given treatment or phenotype should not use models that apply a Poisson distribution, due to the sparsity of the data matrices. Researchers have suggested the use of a negative binomial distribution and log transformation of the data as an alternative to address the over dispersion problem arising in 16S rRNA gene data (Anders and Huber, 2010). However, this tends to increase the false discovery rate due to the data being proportional (Lovell *et al.*, 2015; Gloor *et al.*, 2017).

A recent review suggests alternatives to the standard approaches that have previously been undertaken for each of the major data transformation and analysis steps when dealing with proportional data (Gloor *et al.*, 2017). These include initial normalisation of the count data using a log ratio transformation (centred or isometric) rather than rarefaction. Substitution of beta diversity analysis using Aitchison calculations of distances for Bray Curtis and PHILR for unifracs that beta diversity variance is visualised based on compositional principal component biplots rather than principal co-ordinate. Finally, correlation of proportional data should be performed with an appropriate measure such as ϕ rather than Pearson or Spearman and identifying differential abundant OTUs with ALDEx2 or ANCOM. Methods that account for the compositional aspect of the data have been developed and incorporated into the popular QIIME software package (Caporaso *et al.*, 2010), MixMC and ALDEx2 packages (Fernandes *et al.*, 2014; Mandal *et al.*, 2015; Lê Cao *et al.*, 2016).

Taxonomic identification of OTUs does not define a functional phenotype, as different strains can vary dramatically in their function, that is non-pathogenic *v.* pathogenic. Certain families are more clearly defined and separated, mainly due to the higher level of interest in those groups and therefore more representative sequences are available compared with the less studied families. Analysing data at the family level therefore will generally group many species carrying varied functional capacity together and shed little light on any functional capacity changes to the ecosystem. This also suggests that methods of functional inference such as PICRUSt and Piphillin (Langille *et al.*, 2013; Iwai *et al.*, 2016) although useful, will only be accurate when confidence intervals for a match are high and closest to a reference genome. Efforts to improve this for a rumen microbiome focus have been undertaken through the development of a rumen reference data set, CowPi (Wilkinson *et al.*, 2018).

Applications of microbiota taxonomic profiling

Microbial profiling studies using high throughput sequencing have been performed on the major ruminants investigating changes due to diet, subacute acidosis, methane emissions, feed efficiency, variation along the digestive tract, maternal influence and seasonal changes (Fouts *et al.*, 2012; Lee *et al.*, 2012; Li *et al.*, 2012; Jami *et al.*, 2014; Denman *et al.*, 2015; Mao *et al.*, 2015; Myer *et al.*, 2015; Huws *et al.*, 2016; Martinez-Fernandez *et al.*, 2016; Abecia *et al.*, 2017; Danielsson *et al.*, 2017; Noel *et al.*, 2017; Tapio *et al.*, 2017b;

Wetzels *et al.*, 2017; Petri *et al.*, 2018). But the most comprehensive study performed on ruminates to identify the core microbiota and define what variance in the rumen microbiome was attributed to ruminant species, diet and geographical location was undertaken by the rumen microbial census collaboration (Henderson *et al.*, 2015). A unifying co-ordinated approach limited the large amount of inconsistency in sample preparation, gene amplification and analysis platforms used between different research groups. Diet was found to exhibit the largest influence on the rumen bacterial community, attributed to the changes in the physical and chemical characteristics of the feed allowing for specialised niches. The unclassified Bacteroidales and Ruminococcaceae were observed to be consistently relatively more abundant in animals fed forage, whereas *Prevotella* and unclassified Succinivibrionaceae were relatively more abundant in animals fed diets containing concentrate.

Methanogenic archaeal species were found to be ubiquitously distributed and not affected by host, diet or location with *Methanobrevibacter gottschalkii* and *Methanobrevibacter ruminantium* dominating nearly all samples and accounting for 74% of the archaeal data (Henderson *et al.*, 2015). This reflects the limited substrates used by methanogenic archaea, predominately CO₂ and H₂ or methylated substrates, such as methanol, methylamines and methyl sulphides, and that they exist within the rumen as specialised secondary feeders obtaining substrates for growth from the common fermentative end products of numerous bacterial species.

Several studies including those that have looked at the microbiota populations for natural low methane animals and those that have reduced the methanogen populations through the addition of specific inhibitors have seen common population shifts. Species closely related to the lactic acid producing *Sharpea azabuensis* has been consistently observed to be increased in low methane sheep rumen and likely indicate rapid heterofermentative pathways due to higher rumen turnover rates (Kittelmann *et al.*, 2014; Kamke *et al.*, 2016); whereas H₂ producing Ruminococcaceae, Lachnospiraceae and Verrucomicrobia are more prevalent in high methane yielding animals. In cattle and goats, high methane animals were also associated with increases in Verrucomicrobia and Synergistetes bacteria and decreases in the relative abundance of methanogenic archaea (Denman *et al.*, 2015; Wallace *et al.*, 2015; Martinez-Fernandez *et al.*, 2016). Cattle on high starch diets have also attributed changes in the Succinivibrionaceae relative abundance with methane yields, but this was not evident in cattle on poor quality roughage diets (Martinez-Fernandez *et al.*, 2016; Danielsson *et al.*, 2017).

Prevotella spp. are the most frequently observed OTUs in amplicon data sets and are often both negatively and positively associated with the phenotype being investigated. Yet as most analysis is based on *de novo* clustering, the ability to compare these uncultured diverse OTUs between studies is limited. In a recent review on the ruminal microbiome associated with methane emissions, the authors identified many

Prevotella OTUs from 1000 cattle that clustered together phylogenetically and which fell into groups that were either positively or negatively correlated with methane emissions, suggesting a similar function for closely related species (Tapio *et al.*, 2017a). Further investigation of the functional capacity of these differing but closely related species is needed to provide evidence for their contribution.

Initial reports on a small number of cattle found a positive correlation with methane yield and the bacteria:methanogen ratio (methanogen abundance) (Wallace *et al.*, 2015; Roehe *et al.*, 2016). However, larger groups of animals and other studies have not found this correlation (Danielsson *et al.*, 2012; Wallace *et al.*, 2015; Tapio *et al.*, 2017a). Although the relative abundance of the methanogenic population may not consistently be associated with methane yield and is not informative of the level of functional activity, there is a consistent change in the dominant species with a relative increase in *M. ruminantium* and decrease in *M. Gottschalkii* in low methane animals (Wallace *et al.*, 2015; Danielsson *et al.*, 2017; Martinez-Fernandez *et al.*, 2017).

Microbial profiling for feed efficiency traits in dairy cattle has revealed a reduction in the microbial richness and a strong correlation between the ratio of the phyla Firmicutes to Bacteroidetes and milk-fat yield, but no strong correlation with residual feed intake (RFI) (Jami *et al.*, 2014). However in another study, a similar change in the Firmicutes:Bacteroidetes ratio was evident, but no reduction in the total microbial richness was observed (Myer *et al.*, 2015). However, this may just reflect the different regions of the 16S rRNA gene that were used between the studies, confounding the comparison. Caution should be taken when using Phylum level values and ratios as a phenotype predictor due to the varied functional diversity possessed by the many bacterial species within a phyla. The use of larger numbers of animals and more refined analysis defining distinct microbial composition as an enterotype based on similar microbial clustering should be more accurate (Costea *et al.*, 2018).

Microbial profiling studies using RNA as a template have been used to infer active populations at time of sampling. In a temporal study to monitor fibre colonisation, Huws *et al.* (2016) showed a biphasic colonisation of fibre with primary colonisers established within 1 to 2 h and secondary colonisers not evident until 4 to 8 h after feeding. A constant signal from *Butyrivibrio*, *Fibrobacter*, *Olsenella* and *Prevotella* would indicate their role as core bacteria involved in fibre degradation regardless of the sampling time. The variance in microbes observed due to time after feeding highlights the importance of considering these changes when interpreting the results from a single collection.

Limited studies have focussed on the protozoal and fungal populations of the rumen using amplicon-based methods (Fouts *et al.*, 2012; Kittelmann *et al.*, 2013; Mao *et al.*, 2016; Cunha *et al.*, 2017; Tapio *et al.*, 2017b). Both Kittelmann *et al.* and Cunha *et al.* saw no difference in the fungal populations in relation to methane emissions, whereas Tapio *et al.* indicated that two fungal species were negatively correlated with methane yields (Kittelmann *et al.*, 2014;

Cunha *et al.*, 2017; Tapio *et al.*, 2017a). The taxonomic databases are more poorly characterised for rumen fungi and protozoa, although a specific gut anaerobic fungal database has been created and the 18S rRNA SILVA database is used for protozoa taxonomy (Kittelmann *et al.*, 2012; Yilmaz *et al.*, 2014). A shift towards targeting the D1/D2 region of the 28S rRNA gene for anaerobic fungi should improve classification as an extensively curated taxonomic database exists for the 28S rRNA gene (Edwards *et al.*, 2017).

Other phylogenetic markers targeting functional genes have been used to define specific microbial populations of the rumen, including the formyltetrahydrofolate synthase, methyl coenzyme reductase (*mcrA*) and urease (*ureC*) genes (Gagen *et al.*, 2010; Henderson *et al.*, 2010; Mitsumori *et al.*, 2014; Jin *et al.*, 2017). Only the *ureC* gene has been used in conjunction with high throughput sequencing to study the ureolytic populations of the rumen and rumen mucosal associated populations (Jin *et al.*, 2017). Due to the limited availability of taxonomic data within the reference data set, more than 50% of the data could not be taxonomically identified. However, distinct populations were found on the rumen wall compared with the fibre and liquid associated fractions.

Profiling the rumen microbiota using a taxonomic marker gene is inexpensive, rapid and provides a broad low resolution catalogue of the identified microbiota. Development of robust analysis methods can now accurately correlate shifts in the community and define co-occurrence networks for OTUs, providing greater insight into the complex interactions of the rumen. Initially characterising samples using these methods can aid in deciding which samples are appropriate for more expensive techniques such as metagenomics for describing the functional capacity.

Defining the functional potential of the rumen microbiota through metagenomics analysis

Considerations and limitations

Defining the functional capacity within the rumen microbiota is achieved through sequencing of the combined genomes in a shotgun approach, with the aim of cataloguing the genes and the species to which they belong. Analysis can profile the taxonomy, catalogue the functional genes, attempt to assemble whole genomes and monitor changes in functional gene counts. Metagenomics allows for the study of the uncultivable members and has become an important tool for understanding the full genomic potential that resides within the rumen microbiome, while minimising biases observed with amplicon-based methodologies. Limitations still apply to metagenomics analysis, again such as sample collection and DNA extraction techniques which can bias the proportion and types of species detected. Assembly of sequences into contiguous genomic sequences (contigs) for metagenomics studies are similar to those techniques developed for individual genomes. Although there is a need to overcome issues around varying levels of genomic DNA for different

species (which is exploited by 'binning' techniques, see below) and that closely related species or strains may become inadvertently co-assembled. Most metagenomic assemblers employ varying kmer sizes and coverage depth values to improve *de novo* assembly (IDBA-UD, Meta-IDBA, metaSPAdes and MEGAHIT) (Peng *et al.*, 2011; Bankevich *et al.*, 2012; Peng *et al.*, 2012; Li *et al.*, 2015). Assembled contigs can then be placed into common bins based on nucleotide frequency (most commonly tetramer) and coverage depth within the sample using approaches like PhyloPythiaS, GroopM and MetaBat (Patil *et al.*, 2011; Imelfort *et al.*, 2014; Kang *et al.*, 2015). Sequencing multiple similar environmental samples as opposed to just deeper sequencing of a single sample improves binning based on the assumption that species have the same relative abundance among samples (Albertsen *et al.*, 2013). Completeness and contamination of metagenomic assembled genomes (MAGs) can be assessed based on the presence of multiple lineage-specific single copy marker genes using CheckM or similar approaches with PhyloSift (Darling *et al.*, 2014; Parks *et al.*, 2015). These methodologies also allow for assigning of taxonomy to these MAGs based on these single copy markers being concatenated and placed in concatenated gene trees. Genome-based taxonomy trees are rapidly proving to be more accurate at phylogenetic placement than the single marker gene methods and have recently resulted in 73% of taxa being corrected with one or more changes to their existing taxonomy (Parks *et al.*, 2018).

Applications of metagenomics

Like most other studies of the rumen microbiome, these tools have been employed to gain a better understanding about fibre degradation, methane emissions and ruminant efficiency. Although only performed on three animals and not to the depth of today's metagenomics studies, the first published data defining the fibre adherent population in cattle highlighted the differences obtained from full length PCR-amplified phylogenetic assessment and that from metagenomics data at both the 16S rRNA gene and genomic sequence level (Brulc *et al.*, 2009). Although biases were attributed to the PCR method, all methods exhibited the same power to discriminate between animals and the rumen fluid *v.* fibre adherent populations (Brulc *et al.*, 2009). Assignment of phylogeny for genomic sequences was limited by the relevant data matches in the SEED data set at the time of analysis (Overbeek *et al.*, 2005). However, analysis of the functional capacity concluded that primary colonisers target the easily accessible side chains of complex plant polysaccharides, reflecting that samples were collected 1 h after feeding and that bacteria focussing on the main cellulosic and xylan backbones likely colonise later (Brulc *et al.*, 2009; Huws *et al.*, 2016).

A deeper metagenomic study into the fibre adherent microbiome of switch grass also demonstrated the ability of the rumen microbiota to rapidly colonise and degrade biomass (Hess *et al.*, 2011). Increased sequence data identified 2.5 million open-reading frames (ORFs) of which ~1% were

classified as candidate carbohydrate active genes. The majority of these were novel and not closely aligned with those in the NCBI non-redundant database, highlighting the extensive repertoire of enzymes employed by the rumen microbiota to deconstruct plant material.

Due to the greater sequence data and availability of genome binning methods, the authors were able to group assembled sequence reads based on tetra nucleotide frequencies and read coverage, producing 446 distinct groups (bins). Due to their relatively higher abundance within the fibre adherent population, 15 near complete draft genomes from previously unisolated species were generated (Hess *et al.*, 2011). This allowed for the accurate assignment of functional capacity and potential role to these specific bacteria rather than just gene catalogue counts from the microbiome.

Other ruminants, including Yak and Reindeer have also been investigated, primarily focussing on carbohydrate active enzymes (Dai *et al.*, 2012; Pope *et al.*, 2012). A deep analysis of the Moose metagenome including genomic binning allowed for genomic reconstruction of representatives from the uncultured Bacteroidetes family BS11 and characterisation from genome reconstruction suggested a role in hemicellulosic fermentation (Solden *et al.*, 2017).

Limited sequence depth from a metagenomics study of the dairy rumen microbiome resulted in the assembly of only small contigs and only 20% of these could be functionally annotated (Pitta *et al.*, 2016). However, shifts in the microbiota based on age where detected, along with a high proportion of functional genes assigned to starch degradation which comprised 20% to 30% of the offered diet (Pitta *et al.*, 2016).

Initial studies investigating ruminants in relation to a low methane phenotype, either naturally or through chemical modification did not attempt to generate MAGs, but rather focussed on the phylogenetic and functional changes in the rumen microbiota. Decreases in the relative abundance of methanogenic species and methane generating pathways were evident in goats treated with bromochloromethane (BCM), reflecting the mode of action of BCM on these target species (Denman *et al.*, 2015). Higher H₂ levels in the rumen resulted in fermentative shifts to propionate, which was attributed in the metagenomic data to increases in *Prevotella* and *Selenomonas spp.* and supported by increased functional gene counts for the production of propionate through the randomising (succinate) pathway.

In a cattle trial that selected four pairs of cattle as natural low and high methane emitters, there was a 2.5-fold difference in the archaeal population using qPCR and 16S rRNA relative gene abundance data (Wallace *et al.*, 2015; Roehe *et al.*, 2016). KEGG analysis of archaeal genes associated directly or indirectly with methane production were also higher in the high methane-emitting animals, confirming the increased archaeal presence. A larger group of cattle from varying breeds and diets using a similar analysis also found the same hydrogenotrophic methane synthesis pathway correlated with higher methane emissions and a weak correlation to archaeal abundance (Auffret *et al.*, 2017). However quantitative data for 1000 dairy cows showed only a

weak correlation for the archaea:bacteria ratio with methane emissions (Tapio *et al.*, 2017a), suggesting that the predictive power of this approach may be too low to identify high methane-producing phenotypes.

Similarly, low methane emitting sheep did not show a decrease in the relative abundance of methanogen species or methanogenic pathway genes using metagenomic data, rather only at the gene transcript level was a difference observed (Shi *et al.*, 2014). The bacterial component of the low methane yield sheep rumen metagenome and transcriptome suggested a switch to hexose fermentation through to lactate and butyrate resulting in lower H₂ yields available to drive methanogenesis. This was evident with increased observations of the lactate producing *Sharpea spp.* and subsequent conversion of this substrate by *Megasphaera spp.* to butyrate along with their respective pathways (Kamke *et al.*, 2016).

Metagenomics has also been used to identify functional shifts associated with ruminant efficiency or more commonly low RFI animals (Roehe *et al.*, 2016; Shabat *et al.*, 2016). Efficient dairy cattle defined by RFI exhibited a lower richness of abundant microbial species for both 16S rRNA and at the microbial functional gene level (Shabat *et al.*, 2016). Both phylogenetically and functionally efficient animals were dominated by increases in *Megasphaera elsdenii* and *Coprococcus catus*, hydrogen consuming lactate utilisers resulting in production of butyrate and propionate for the host. *Coprococcus spp.* have also previously been associated with the NADPH-dependent reduction of phloroglucinol and the redirection of H₂ in the rumen to acetate in a methane inhibited rumen (Martinez-Fernandez *et al.*, 2017). In dairy cattle, efficient animals were also linked with decreases in *Methanobrevobacter spp.* and methanogenic pathway genes (Shabat *et al.*, 2016). Thus, the microbiology associated with H₂ production and utilisation within the rumen seems to be tightly linked with methane and efficiency traits.

In cattle that were initially selected for their methane emission ranking, a correlation with RFI was also detected in which the authors demonstrated the abundance of 49 genes, explaining 86% of the variation observed in feed efficiency (Roehe *et al.*, 2016). Of particular note were genes identified as 'fucose sensing' involved in cross talk between the host and the microbiota, possibly in response to the mucin content from bovine saliva.

Recently, results from the Hungate 1000 genome project were published, revealing genomic coverage of ~75% of the known genera from the rumen (Seshadri *et al.*, 2018). But in spite of that the Hungate collection covers only a fraction of the diversity found within the rumen (Li *et al.*, 2018; Stewart *et al.*, 2018). Close to 2.2% of the ORFs in the combined genomes were classified as carbohydrate-active enzymes and binding proteins reflecting the functional role of bacteria in degrading cellulose, hemicellulose and pectin. Furthermore, polysaccharide utilisation loci involved in the degradation of animal glycans were enriched in the Bacteroidetes genomes and may indicate the ability of these species to harvest energy from *N*-linked salivary glycoproteins (Seshadri

et al., 2018). Metabolic fermentation pathway reconstruction for the sequenced species has now lead to the most complete reconstruction of the rumen microbiome metabolic scheme incorporating both species and functional capacity. However, the level of contribution that these species make to the complex interactions that take place within the rumen is still not fully understood. Although, recently the rumen microbiome gene catalogue clearly observed diet modulation in gene abundance counts even though 90% of coding genes were shared (Li *et al.*, 2018).

Likewise, several studies focussing on the generation of MAGs from ruminants have discovered similar over representation of carbohydrate active enzymes for 99, 324 and 913 MAGs from moose and cattle (Svartstrom *et al.*, 2017; Li *et al.*, 2018; Stewart *et al.*, 2018).

With the results from the Hungate 1000 genome sequencing initiative and the various rumen focussed MAGs and rumen microbiota gene catalogue available, there is the future possibility of reducing the requirement to perform metagenomic assemblies and rather rapidly generate relative abundance counts and functional capacity linked to taxonomy through direct mapping to rumen relevant annotated genomic data sets (Hess *et al.*, 2011; Parks *et al.*, 2017; Svartstrom *et al.*, 2017; Li *et al.*, 2018; Seshadri *et al.*, 2018; Stewart *et al.*, 2018).

As of yet very little genomic data representing protists and the anaerobic fungi are available, substantially limiting our ability to study these groups. However, improvements in long read sequencing platforms have overcome some of the difficulties in assembling the highly repetitive, AT base rich genomes of the anaerobic fungi (Solomon *et al.*, 2016; Haitjema *et al.*, 2017). Comparative genomics and proteomics have catalogued an extensive array of plant depolymerisation and structural genes that form the anaerobic fungal cellulosome complex. Scaffoldin proteins, containing dokerin-binding cohesion sequences were highly conserved across the Neocallimastigomycota, allowing potentially for interspecies fungal enzyme complexes to form (Haitjema *et al.*, 2017). This large diversity of degradative and substrate binding capacity incorporated into the fungal cellulosome structure provides a co-ordinated and synergistic mechanism for complete conversion of cellulosic material to fermentable sugars. Transcriptomics provided further insights into aspects of the co-ordination, identifying repressive regulation of cellulosic gene sets in response to glucose as an end product metabolite. Furthermore, the tailoring of hydrolytic gene transcripts was linked to the complexity of the substrate through an increase in the number and functional diversity of degradative genes transcribed (Solomon *et al.*, 2016).

Observing functional gene activity within the rumen using metatranscriptomics

Considerations and limitations

Metagenomic analysis performed on genomic DNA cannot distinguish whether the material is from viable cells or if the

predicted genes are functionally expressed at the time of collection. However, limitations for using rRNA as an indicator of the species activity within a community exist and should be considered. Concentrations of rRNA are not consistently correlated with growth and can differ greatly between closely related taxa, whereas dormant cells can still contain high levels of rRNA (Blazewicz *et al.*, 2013). Unlike eukaryotic mRNA, the majority of prokaryotic transcripts are not poly adenylated at the 3' end, making the commonly used mRNA polyA enrichment methods impractical (Sarkar, 1997). Most current microbial targeted methods employ techniques to deplete the rRNA sequences in order to increase the number of non-rRNA reads in the data sets.

Computational methods to identify and remove the ribosomal sequences are common and easily employed. Likewise host transcripts should be removed if a reference data set is available in order to enrich the microbiome transcripts. Although, host transcript data should not be discarded as it may show some indication of the cross talk between the host and its microbiome. Due to the lack of relevant rumen microbial genomic data sets to map transcripts too, most analyses involve a *de novo* assembly and annotation step.

Application of metatranscriptomics

Considering that most RNASeq analysis workflows were established for eukaryotic mRNA, it was not surprising that the first large-scale rumen metatranscriptomic analysis focussed on the rumen eukaryotic species. Even though samples were collected before feeding, the eukaryotic fibre adherent population of the muskoxen rumen exhibited a high level of expression for genes involved in the degradation of crystalline plant cell wall polysaccharides, with large numbers of exo-acting glucanases and swollenin genes (Qi *et al.*, 2011). Functional assignment of the transcripts showed that 3.4% of the data were more closely related to bacterial sequences, illustrating the poor coverage of rumen eukaryotic sequences in the reference database and the occurrence of horizontal gene transfer that occurs within the rumen microbiome.

Using methods to deplete the ribosomal RNA sequences from the total RNA, Dai and colleagues were able to collect in excess of 1 million non-rRNA reads of which ~1% were identified as carbohydrate active enzymes or binding modules (Dai *et al.*, 2015). A similar level of carbohydrate active genes were observed in the mRNA enriched metatranscriptomic data of dairy cows in Japan (Shinkai *et al.*, 2016). These studies confirmed that the major bacterial activity of fibre degradation was being performed by members of the genera *Fibrobacter*, *Prevotella* and *Ruminococcus*. Recently, a study on dairy cows in France used 18 newly designed ribosomal depletion capture probes covering a large number of the rumen archaeal, bacterial, fungal and protozoal genera to enrich for non-rRNA reads (Comtet-Marre *et al.*, 2017). Likewise, the data confirmed the majority of bacterial activity resides with these fibrolytic species, but also highlighted the large contribution from fungal and protozoal species. The lack of relevant genomes for mapping and annotation of

RNA-Seq data is still considered the major limitation of these methods.

Metatranscriptomic analysis of low and high RFI beef cattle used total RNA sequencing rather than enriching for non-rRNA sequences (Li and Guan, 2017). There was no variance between the microbiomes based on the rRNA data at the taxonomic family level. Approximately 90% of all the read data were identified as rRNA, resulting on an average ~4.5 million non-ribosomal reads for transcriptomic analysis. Based on reads assigned to KEGG pathways, 30 pathways were relatively more abundant in H-RFI animals (inefficient), including pathways associated with amino acid metabolism, reinforcing the findings from H-RFI dairy cows using metagenomics (Shabat *et al.*, 2016).

Early studies of rumen methanogens using rRNA and *mcrA* libraries identified a group of archaea closely related to Thermoplasmatales (rumen cluster C) (Denman *et al.*, 2007; Janssen and Kirs, 2008). Although it was hypothesised that they were likely to be methanogenic archaea, due the ubiquitous nature of this group within the rumen and supported by the presence of the *mcrA* gene. It was not until Poulsen and colleagues used metatranscriptomics that a definitive link could be ascertained between methanogenic capacity and the RCC group (Poulsen *et al.*, 2013). Based on these transcripts, it was concluded that this group of methanogens used methylamines for the production of methane rather than the usual hydrogenotrophic pathway used by *Methanobrevibacter* spp. The RCC and related methylotrophic methanogens were renamed and belong to the new order Methanomassiliicoccales (Iino *et al.*, 2013).

High methane-emitting sheep did not show any variance in the abundance of methanogenic species or pathways from metagenomic analysis, but did show a strong correlation between hydrogenotrophic methanogenesis transcripts and methane yields (Shi *et al.*, 2014). It was concluded that this was likely a response to the supply of hydrogen in the rumen from the fermentative processes of other rumen microbes. Further to this study, the bacterial species were the focus of a second paper in which it was suggested that due most likely to a smaller rumen size and faster transit times, conditions were more favourable for rapid bacterial fermentation of hexose through to lactate and butyrate resulting in lower H₂ yields for methanogenesis (Kamke *et al.*, 2016). Butyrate production was assigned to *Sharpea* spp. and *Megasphaera* spp. and involves a two-step process, calculated to produce 24% less methane, due to less net hydrogen generation, compared with the common one-step fermentative pathway performed by Ruminococcaceae (Kamke *et al.*, 2016).

Metaproteomics for the characterisation of the rumen microbioata proteome

Considerations and limitations

Advances in the throughput and accuracy of mass spectrometry tools coupled with peptide separation methods can now lead to the detection of over 100 000 tandem mass

spectra per acquisition (Timmins-Schiffman *et al.*, 2017). But like all other omic approaches, the limitation for identification and taxonomic assignment is due to the poor coverage of sample relevant data in the reference sets. Workflows that combine relevant metagenomic and isolate sequence data will improve peptide identification (Petritz and Franco, 2017). Processing methods that reduce contaminating proteins from the diet or host are also essential to maximise the focus on the microbiome peptides. Likewise, rumen and faecal samples can contain polyphenolic compounds, such as tannins and humic acids that can co-precipitate and interfere with analysis through modifications that make it difficult to identify the peptides and suppress ionisation (Makkar *et al.*, 1995; Qian and Hettich, 2017; Snelling and Wallace, 2017).

Applications of metaproteomics

Limited metaproteomic studies have been published for the rumen, an early attempt from sheep rumen microorganisms focussed on identifying cellulose-binding proteins through an enrichment step (Toyoda *et al.*, 2009). By using 1D-PAGE coupled to MS/MS, they were able to identify a small number of proteins and link them to microbial species using the limited databases available at the time. Endoglucanases of the cellulolytic bacterium *Fibrobacter succinogenes* and an exoglucanase from the fungi *Piromyces equi* were among the proteins identified.

Increased resolution and substantially more peptides were identified when 2D-PAGE separation was used before LC-MS/MS detection (Snelling and Wallace, 2017). Issues associated with humic content of grass fed animals and between-sample replication limits the possibility to use 2D-PAGE as a sole tool to predict the function of rumen proteins. Humic compounds co-precipitate with proteins and alter gel mobility, resulting in unresolved smears rather than distinct protein bands. However, metaproteomics of the separated proteins identified protozoal structural proteins, prokaryotic central metabolic enzymes and archaeal methanogenesis proteins (Snelling and Wallace, 2017).

Using a metaproteomic shotgun approach to discover peptides from plant adherent and rumen liquid fraction microbiota, it was possible to identify in excess of 2000 bacterial, 150 archaeal and 800 eukaryotic proteins in the fibre adherent fraction and similar ratios but lower numbers in the liquid fraction (Deusch and Seifert, 2015). Bacterial and archaeal taxonomy was as expected with Prevotellaceae, Fibrobacteraceae, Ruminococcaceae, Clostridiaceae, Methanobacteriaceae and Methanomicrobiaceae dominating. Eukaryotic taxonomic assignment was hindered by the sparsity of rumen protozoal and fungal data sets and most proteins could only be identified to the Phyla level and were classified as originating from plant and host proteins.

A more detailed metaproteomic study including sample fractionation for fibre associated and liquid microbiota, along with diet shifts from forage and grain based diets produced over 8000 bacterial and 350 archaeal proteins (Deusch *et al.*, 2017). Concurrent with the 16S rRNA amplicon-based analysis of the same samples, diet was

confirmed as the largest driver of microbiota change. Succinivibrionaceae OTUs and proteins attributed to this group, particularly carbohydrate esterases were both more abundant in the corn-supplemented diet and confirmed its increase on starch-based diets as observed previously (Petri *et al.*, 2013; Wallace *et al.*, 2015; Martinez-Fernandez *et al.*, 2016). The major fibre degrading bacteria *Fibrobacter spp.*, *Ruminococcus spp.* and Lachnospiraceae were more prevalent in the solid fraction. Furthermore, a high proportion of proteins linked to butyrate formation were assigned to Lachnospiraceae bacteria in the fibre-rich diets. The metabolically diverse Prevotellaceae were abundant in all samples and linked with acetate and propionate pathway proteins (Deusch *et al.*, 2017). Archaeal populations did not differ significantly between diets or collected fractions due to low abundance. However, proteomic analysis identified proteins from the Crenarchaeota and Thaumarchaeota Phyla that the amplicon sequencing did not, possibly because of primer biases for the OTU sequencing. The methyl-coenzyme M reductase involved in the final step of methanogenesis was detected at its lowest in the corn-based diet, corresponding with the general observation of others that methanogenesis is lower in cattle fed high concentrate diets (Rooke *et al.*, 2014; Martinez-Fernandez *et al.*, 2016).

Microbial metabolite detection within the rumen using metabolomics

The majority of rumen nutrition studies have usually performed some limited metabolomics analysis, at least presenting data on short chain fatty acids (SCFA) products and some with methane and hydrogen levels. Mass spectrometry and nuclear magnetic resonance spectroscopy are the most popular high throughput methods being applied to rumen samples.

Analogous to most other rumen analysis, the metabolome of animals on different diets could easily be distinguished. Most metabolomics studies in ruminants have investigated the increased proportion of concentrate/grain in a diet compared with roughage. Saleem *et al.* (2013) summarised the results of studies performed on increasing levels of concentrate using various methods to detect 246 rumen metabolites, covering phospholipids, inorganic ions, gases, amino acids SCFA and carbohydrates (Ametaj *et al.*, 2010; Saleem *et al.*, 2012). Increased grain in the diet lead to rises in methylamines and a decrease in 3-phenylpropionate (hydrocinnamate), likely linked to the decreased plant phenolic component of the diet. Similar findings were also reported from other studies when altering the diet (Zhao *et al.*, 2014; Mao *et al.*, 2016; Zhang *et al.*, 2017a and 2017b; O'Callaghan *et al.*, 2018).

Profiling of feed efficient rumens could also be distinguished by metabolomics, ruminal biohydrogenation pathways, in particular down regulation of linoleic and alpha linoleic acid metabolism were associated with average daily gain (Artegoitia *et al.*, 2017). However, feed efficiency in

dairy cows was linked to increased production of SCFA, with increased propionate:acetate ratios and decreased methane and increased putrescine (Shabat *et al.*, 2016).

Unlike other omic technologies, metabolomics cannot directly link metabolites to a microbial species. Associations with changes in microbial relative abundance through microbial profiling, metatranscriptomic or metaproteomic is required. Correlations between changes in urine and plasma metabolites have been linked to relative abundance shifts for protozoa and *Methanomassiliicoccus spp.*, particularly trimethylamine *N*-oxide (Morgavi *et al.*, 2015; Saro *et al.*, 2018). This suggests that this compound may be used as a biomarker for monitoring methylamine utilising methanogens through urine metabolites.

Combining many of the omic platforms will often lead to more powerful observations; for determining the function of un-cultivated bacteria or confirming functional roles within different systems. The characterisation of multiple genomes from the un-cultivated Bacteroidetes BS11 family using metagenomic data and confirmation of function with metaproteomic and metabolomic data categorised them as hemicellulosic degraders producing SCFA as end products (Solden *et al.*, 2017). Likewise, microbial profiling partnered with metaproteomics and metabolomics reveals the Prevotellaceae as a metabolically versatile group dominating on both concentrate and fibre rich diets (Deusch *et al.*, 2017). Corn-based diets promoted the activities of the Succinivibrionaceae family, leading to production of succinate for use by other species resulting in propionate as an end product. Fibre rich diets promoted Prevotellaceae, *Ruminococcus spp.* and Lachnospiraceae to drive butyrate and propionate formation leading to changes in the propionate:acetate ratio. Dairy cattle showed increased energy harvest in feed efficient animals that was linked to a restriction in microbiome diversity and richness. Efficient animals were highlighted with increases in hydrogen and lactate utilising *Megasphaera elsdenii* and *Coprococcus catus*, resulting in production of butyrate and propionate (Shabat *et al.*, 2016).

Future directions

Accurate annotation of the functional gene products allows for the construction of genome-scale metabolic models (GEM) for individual isolates, which can then be applied to understand complex interactions within the system (van der Ark *et al.*, 2017). Most microbes have the capacity to utilise a wide array of nutrients using varied metabolic pathways. Tools such as Minimal Environmental TOol (MENTO), RAVEN and ModelSEED can be used to predict the minimal nutrient requirements of as yet un-cultivated organisms based on its GEM (Henry *et al.*, 2010; Agren *et al.*, 2013; Zarecki *et al.*, 2014). Beyond models that predict an individual's phenotype are ones that allow for modelling co-cultures and multi-species interactions to predict ways to drive a desired rumen

phenotype. Hypothesis testing based on culturing studies can be performed on isolates and multiple species competing and interacting to build network models to better inform the GEM in an iterative manner. Predictive competitive and cooperative metabolic models already suggest that competition is generally dominated by versatile fast growing species (Freilich *et al.*, 2011). Likely explaining the high abundance of *Prevotella spp.* consistently observed in the rumen. Regardless of these techniques, accuracy and further advancement of these models will still require isolation and culturing of these new microorganisms to validate and strengthen predictions. New high throughput culturing methods and media are proving successful (Kenters *et al.*, 2011; Lagier *et al.*, 2016).

Monitoring changes to rumen function for research purposes is achievable, but scaling to the level that will drive changes in production systems will need the development of simple robust quantitative markers. A large number of microbial metabolites and microbial-host co-metabolites are present in plasma and other body fluids. There is potential for the discovery of biomarkers linked to microbiota function that could be applied in farms using less invasive and simpler techniques. Saliva and buccal swab samples have already been shown to reflect the rumen microbiome, whereas changes in plasma fatty acid profiles were suggested as biomarkers for weight gain and levels of trimethylamine *N*-oxide as a marker for methylamine utilising methanogens (Kittelman *et al.*, 2015; Morgavi *et al.*, 2015; Tapio *et al.*, 2016; Artegoitia *et al.*, 2017).

The omics technologies have given us an understanding of which species are present and their potential function. Now, why and how they are present will start to be addressed. Ultimately, these understandings should lead to the ability to design targeted interventions that direct rumen composition and activity towards improved production, health and benefits to the environment.

Conclusions

Considering the limitations and erroneous methods initially used, primarily due to the infancy of these technologies, the low numbers of animals investigated and lack of robust statistical tools for compositional data, omics based technologies have produced results that relate to and support our biological understanding of the rumen system. Continued and more precise use of these tools will lead to a detailed and comprehensive understanding of compositional and functional capacity. Relative abundance shifts in microbial populations are now being related to gene transcripts and proteins that explain changes in detected metabolites. However, our ability to rationally design and drive rumen microbial composition and function in this highly complex and dynamically changing environment is limited. Integrating omics data will allow for the construction of rumen-specific microbial metabolic models.

Acknowledgements

D. P. Morgavi was partially supported by the French National Research Agency (ANR) through the FACCE-JPI project RumenStability and FACCE ERA-GAS project RumenPredict. S. E. Denman and C. S. McSweeney were partially supported by Meat and Livestock Australia (MLA).

Declaration of interest

None.

Ethics statement

All ethical standards have been met.

Software and data repository resources

None.

References

- Abecia L, Jiménez E, Martínez-Fernández G, Martín-García AI, Ramos-Morales E, Pinloche E, Denman SE, Newbold CJ and Yáñez-Ruiz DR 2017. Natural and artificial feeding management before weaning promote different rumen microbial colonization but not differences in gene expression levels at the rumen epithelium of newborn goats. *PLoS One* 12, e0182235.
- Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I and Nielsen J 2013. The RAVEN toolbox and its use for generating a genome-scale metabolic model for *penicillium chrysogenum*. *PLoS Computational Biology* 9, e1002980.
- Aitchison J, Barceló-Vidal C, Martín-Fernández JA and Pawłowsky-Glahn V 2000. Logratio analysis and compositional distance. *Mathematical Geology* 32, 271–275.
- Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW and Nielsen PH 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology* 31, 533–538.
- Ametaj BN, Zebeli Q, Saleem F, Psychogios N, Lewis MJ, Dunn SM, Xia J and Wishart DS 2010. Metabolomics reveals unhealthy alterations in rumen metabolism with increased proportion of cereal grain in the diet of dairy cows. *Metabolomics* 6, 583–594.
- Anders S and Huber W 2010. Differential expression analysis for sequence count data. *Genome Biology* 11, R106.
- Artegoitia VM, Foote AP, Lewis RM and Freetly HC 2017. Rumen fluid metabolomics analysis associated with feed efficiency on crossbred steers. *Scientific Reports* 7, 2864.
- Auffret MD, Stewart R, Dewhurst RJ, Duthie CA, Rooke JA, Wallace RJ, Freeman TC, Snelling TJ, Watson M and Roehe R 2017. 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.
- Balvočiūtė M and Huson DH 2017. SILVA, RDP, Greengenes, NCBI and OTT — how do these taxonomies compare? *BMC Genomics* 18, 114.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA and Pevzner PA 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19, 455–477.
- Blazewicz SJ, Barnard RL, Daly RA and Firestone MK 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *The ISME Journal* 7, 2061–2068.
- Bragg L, Stone G, Imelfort M, Hugenholtz P and Tyson GW 2012. Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nature Methods* 9, 425–426.
- Brucl JM, Antonopoulos DA, Miller ME, Wilson MK, Yannarell AC, Dinsdale EA, Edwards RE, Frank ED, Emerson JB, Wacklin P, Coutinho PM, Henrissat B, Nelson KE and White BA 2009. Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences of the United States of America* 106, 1948–1953.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA and Holmes SP 2016. DADA2: high-resolution sample inference from illumina amplicon data. *Nature Methods* 13, 581–583.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J and Knight R 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7, 335–336.
- Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, Chandra S, McGarrell DM, Schmidt TM, Garrity GM and Tiedje JM 2003. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Research* 31, 442–443.
- Comtet-Marre S, Parisot N, Lepercq P, Chaucheyras-Durand F, Mosoni P, Peyretaillade E, Bayat AR, Shingfield KJ, Peyret P and Forano E 2017. Metatranscriptomics reveals the active bacterial and eukaryotic fibrolytic communities in the rumen of dairy cow fed a mixed diet. *Frontiers in Microbiology* 8, 67.
- Costea PI, Hildebrand F, Arumugam M, Bäckhed F, Blaser MJ, Bushman FD, de Vos WM, Ehrlich SD, Fraser CM, Hattori M, Huttenhower C, Jeffery IB, Knights D, Lewis JD, Ley RE, Ochman H, O'Toole PW, Quince C, Relman DA, Shanahan F, Sunagawa S, Wang J, Weinstock GM, Wu GD, Zeller G, Zhao L, Raes J, Knight R and Bork P 2018. Enterotypes in the landscape of gut microbial community composition. *Nature Microbiology* 3, 8–16.
- Cunha CS, Veloso CM, Marcondes MI, Mantovani HC, Tomich TR, Pereira LGR, Ferreira MFL, Dill-McFarland KA and Suen G 2017. Assessing the impact of rumen microbial communities on methane emissions and production traits in Holstein cows in a tropical climate. *Systematic and Applied Microbiology* 40, 492–499.
- Dai X, Tian Y, Li J, Su X, Wang X, Zhao S, Liu L, Luo Y, Liu D, Zheng H, Wang J, Dong Z, Hu S and Huang L 2015. Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in the cow rumen. *Applied and Environmental Microbiology* 81, 1375–1386.
- Dai X, Zhu Y, Luo Y, Song L, Liu D, Liu L, Chen F, Wang M, Li J, Zeng X, Dong Z, Hu S, Li L, Xu J, Huang L and Dong X 2012. Metagenomic insights into the fibrolytic microbiome in yak rumen. *PLoS One* 7, e40430.
- Danielsson R, Dicksved J, Sun L, Gonda H, Müller B, Schnürer A and Bertilsson J 2017. Methane production in dairy cows correlates with rumen methanogenic and bacterial community structure. *Frontiers in Microbiology* 8, 226.
- Danielsson R, Schnürer A, Arthurson V and Bertilsson J 2012. Methanogenic population and CH₄ production in Swedish dairy cows fed different levels of forage. *Applied and Environmental Microbiology* 78, 6172–6179.
- Darling AE, Jospin G, Lowe E, Matsen FAIV, Bik HM and Eisen JA 2014. PhyloSift: phylogenetic analysis of genomes and metagenomes. *Peer J* 2, e243.
- Denman SE, Martínez-Fernández G, Shinkai T, Mitsumori M and McSweeney CS 2015. Metagenomic analysis of the rumen microbial community following inhibition of methane formation by a halogenated methane analog. *Frontiers in Microbiology* 6, 1087.
- Denman SE, Tomkins NW and McSweeney CS 2007. Quantitation and diversity analysis of ruminal methanogenic populations in response to the anti-methanogenic compound bromochloromethane. *FEMS Microbiology Ecology* 62, 313–322.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P and Andersen GL 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* 72, 5069–5072.
- Deusch S, Camarina-Silva A, Conrad J, Beifuss U, Rodehutschord M and Seifert J 2017. A structural and functional elucidation of the rumen microbiome influenced by various diets and microenvironments. *Frontiers in Microbiology* 8, 1605.
- Deusch S and Seifert J 2015. Catching the tip of the iceberg – evaluation of sample preparation protocols for metaproteomic studies of the rumen microbiota. *Proteomics* 15, 3590–3595.
- Dore J and Stahl DA 1991. Phylogeny of anaerobic rumen chytridiomycetes inferred from small subunit ribosomal-rna sequence comparisons. *Canadian Journal of Botany-Revue Canadienne De Botanique* 69, 1964–1971.
- Edgar R 2018. Taxonomy annotation and guide tree errors in 16S rRNA databases. *PeerJ* 6, e5030.
- Edwards JE, Forster RJ, Callaghan TM, Dollhofer V, Dagar SS, Cheng Y, Chang J, Kittelmann S, Fliegerova K, Puniya AK, Henske JK, Gilmore SP, O'Malley MA,

- Griffith GW and Smidt H 2017. PCR and omics based techniques to study the diversity, ecology and biology of anaerobic fungi: insights, challenges and opportunities. *Frontiers in Microbiology* 8, 1657.
- Fernandes AD, Reid JN, Macklaim JM, McMurry TA, Edgell DR and Gloor GB 2014. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* 2, 15.
- Fouts DE, Szpakowski S, Purushe J, Torralba M, Waterman RC, MacNeil MD, Alexander LJ and Nelson KE 2012. Next generation sequencing to define prokaryotic and fungal diversity in the bovine rumen. *PLoS One* 7, e48289.
- Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M, Gophna U, Sharan R and Ruppin E 2011. Competitive and cooperative metabolic interactions in bacterial communities. *Nature Communications* 2, 589.
- Gagen EJ, Denman SE, Padmanabha J, Zadbuke S, Al Jassim R, Morrison M and McSweeney CS 2010. Functional gene analysis suggests different acetogen populations in the bovine rumen and tammar wallaby forestomach. *Applied and Environmental Microbiology* 76, 7785–7795.
- Gantner S, Andersson AF, Alonso-Saez L and Bertilsson S 2011. Novel primers for 16S rRNA-based archaeal community analyses in environmental samples. *Journal of Microbiological Methods* 84, 12–18.
- Gloor GB, Macklaim JM, Pawlowsky-Glahn V and Egozcue JJ 2017. Microbiome datasets are compositional: and this is not optional. *Frontiers in Microbiology* 8, 2224.
- Haitjema CH, Gilmore SP, Henske JK, Solomon KV, de Groot R, Kuo A, Mondo SJ, Salamov AA, LaButti K, Zhao Z, Chiniquy J, Barry K, Brewer HM, Purvine SO, Wright AT, Hainaut M, Boxma B, van Alen T, Hackstein JHP, Henrissat B, Baker SE, Grigoriev IV and O'Malley MA 2017. A parts list for fungal cellulosesomes revealed by comparative genomics. *Nature Microbiology* 2, 17087.
- Henderson G, Cox F, Ganesh S, Jonker A, Young W and Janssen PH 2015. Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Scientific Reports* 5, 14567.
- Henderson G, Cox F, Kittelmann S, Miri VH, Zethof M, Noel SJ, Waghorn GC and Janssen PH 2013. Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLoS One* 8, e74787.
- Henderson G, Naylor GE, Leahy SC and Janssen PH 2010. Analysis of formyltetrahydrofolate synthetase sequences from ruminants reveals the presence of new potentially homoacetogenic bacteria in the rumen. *Applied Environmental Microbiology* 76, 2058–2066.
- Henry CS, DeJongh M, Best AA, Frybarger PM, Lindsay B and Stevens RL 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology* 28, 977–982.
- Hess M, Sczyrba A, Egan R, Kim TW, Chokhwalwa H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z and Rubin EM 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* 331, 463–467.
- Huws SA, Edwards JE, Creevey CJ, Rees Stevens P, Lin W, Girdwood SE, Pachebat JA and Kingston-Smith AH 2016. Temporal dynamics of the metabolically active rumen bacteria colonizing fresh perennial ryegrass. *FEMS Microbiology Ecology* 92, fiv137.
- Iino T, Tamaki H, Tamazawa S, Ueno Y, Ohkuma M, Suzuki K, Igarashi Y and Haruta S 2013. *Candidatus Methanogranum caenicola*: a novel methanogen from the anaerobic digested sludge, and proposal of *Methanomassiliicoccaceae* fam. nov. and *Methanomassiliicoccales* ord. nov., for a methanogenic lineage of the class Thermoplasmata. *Microbes and Environments* 28, 244–250.
- Imelfort M, Parks D, Woodcroft BJ, Dennis P, Hugenholtz P and Tyson GW 2014. GroopM: an automated tool for the recovery of population genomes from related metagenomes. *PeerJ PrePrints* 2, e409v401.
- Iwai S, Weinmaier T, Schmidt BL, Albertson DG, Poloso NJ, Dabbagh K and DeSantis TZ 2016. Piphillin: improved prediction of metagenomic content by direct inference from human microbiomes. *PLoS One* 11, e0166104.
- Jami E, White BA and Mizrahi I 2014. Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. *PLoS One* 9, e85423.
- Janssen PH and Kirs M 2008. Structure of the archaeal community of the rumen. *Applied Environmental Microbiology* 74, 3619–3625.
- Jin D, Zhao S, Zheng N, Bu D, Beckers Y, Denman SE, McSweeney CS and Wang J 2017. Differences in ureolytic bacterial composition between the rumen digesta and rumen wall based on ureC gene classification. *Frontiers in Microbiology* 8, 385.
- Kamke J, Kittelmann S, Soni P, Li Y, Tavendale M, Ganesh S, Janssen PH, Shi W, Froula J, Rubin EM and Attwood GT 2016. Rumen metagenome and meta-transcriptome analyses of low methane yield sheep reveals a *Sharpea*-enriched microbiome characterised by lactic acid formation and utilisation. *Microbiome* 4, 56.
- Kang DD, Froula J, Egan R and Wang Z 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3, e1165.
- Kenters N, Henderson G, Jeyanathan J, Kittelmann S and Janssen PH 2011. Isolation of previously uncultured rumen bacteria by dilution to extinction using a new liquid culture medium. *Journal of Microbiological Methods* 84, 52–60.
- Kittelmann S, Kirk MR, Jonker A, McCulloch A and Janssen PH 2015. Buccal swabbing as a non-invasive method to determine bacterial, archaeal, and eukaryotic microbial community structure in the rumen. *Applied and Environmental Microbiology* 81, 7470–7483.
- Kittelmann S, Naylor GE, Koolaard JP and Janssen PH 2012. A proposed taxonomy of anaerobic fungi (class neocallimastigomycetes) suitable for large-scale sequence-based community structure analysis. *PLoS One* 7, e36866.
- Kittelmann S, Pinares-Patiño CS, Seedorf H, Kirk MR, Ganesh S, McEwan JC and Janssen PH 2014. Two different bacterial community types are linked with the low-methane emission trait in sheep. *PLoS One* 9, e103171.
- Kittelmann S, Seedorf H, Walters WA, Clemente JC, Knight R, Gordon JI and Janssen PH 2013. Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *PLoS One* 8, e74789.
- Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M and Glöckner FO 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research* 41, e1.
- Konstantinidis KT and Tiedje JM 2007. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Current Opinion in Microbiology* 10, 504–509.
- Kozich JJ, Westcott SL, Baxter NT, Highlander SK and Schloss PD 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied Environmental Microbiology* 79, 5112–5120.
- Kunin V, Engelbrekton A, Ochman H and Hugenholtz P 2009. Wrinkles in the rare biosphere: pyrosequencing errors lead to artificial inflation of diversity estimates. *Environmental Microbiology* 12, 118–123.
- Lagier J-C, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, Caputo A, Cadoret F, Traore SI, Seck EH, Dubourg G, Durand G, Mourembou G, Guilhot E, Togo A, Bellali S, Bachar D, Cassir N, Bittar F, Delerac J, Mailhe M, Ricaboni D, Bilen M, Dangui Niekro NPM, Dia Badiane NM, Valles C, Mouelhi D, Diop K, Million M, Musso D, Abrahão J, Azhar EI, Bibi F, Yasir M, Diallo A, Sokhna C, Djossou F, Vitton V, Robert C, Rolain JM, La Scola B, Fournier P-E, Levasseur A and Raoult D 2016. Culture of previously uncultured members of the human gut microbiota by culturomics. *Nature Microbiology* 1, 16203.
- Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkpile DE, Vega Thurber RL, Knight R, Beiko RG and Huttenhower C 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology* 31, 814–821.
- Lê Cao K-A, Costello M-E, Lakis VA, Bartolo F, Chua X-Y, Brazeilles R and Rondeau P 2016. MixMC: a multivariate statistical framework to gain insight into microbial communities. *PLoS One* 11, e0160169.
- Lee HJ, Jung JY, Oh YK, Lee SS, Madsen EL and Jeon CO 2012. Comparative survey of rumen microbial communities and metabolites across one caprine and three bovine groups, using bar-coded pyrosequencing and (1)H nuclear magnetic resonance spectroscopy. *Applied and Environmental Microbiology* 78, 5983–5993.
- Li D, Liu CM, Luo R, Sadakane K and Lam TW 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676.
- Li F and Guan LL 2017. Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle. *Applied and Environmental Microbiology* 83, e00061–e00017.
- Li J, Zhong H, Ramayo-Caldas Y, Terrapon N, Lombard V, Potocki-Veronese G, Estelle-Fabrellas J, Popova M, Yang Z, Zhang H, Li F, Tang S, Chen W, Chen B, Li

- J, Guo J, Martin C, Maguin E, Xu X, Yang H, Wang J, Madsen L, Kristiansen K, Henrrisat B, Ehrlich SD and Morgavi DP 2018. A catalog of microbial genes from the bovine rumen reveals the determinants of herbivory. *bioRxiv* 272690. <https://doi.org/10.1101/272690>.
- Li RW, Connor EE, Li C, Baldwin Vi RL and Sparks ME 2012. Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools. *Environmental Microbiology* 14, 129–139.
- Louca S, Doebeli M and Parfrey LW 2018. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome* 6, 41.
- Lovell D, Pawlowsky-Glahn V, Egozcue JJ, Marguerat S and Bahler J 2015. Proportionality: a valid alternative to correlation for relative data. *PLoS Computational Biology* 11, e1004075.
- Makkar HP, Blummel M and Becker K 1995. Formation of complexes between polyvinyl pyrrolidones or polyethylene glycols and tannins, and their implication in gas production and true digestibility in *in vitro* techniques. *British Journal of Nutrition* 73, 897–913.
- Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R and Peddada SD 2015. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology in Health and Disease* 26, 10.3402/mehd.v3426.27663.
- Mao S, Zhang M, Liu J and Zhu W 2015. Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: membership and potential function. *Scientific Reports* 5, 16116.
- Mao SY, Huo WJ and Zhu WY 2016. Microbiome-metabolome analysis reveals unhealthy alterations in the composition and metabolism of ruminal microbiota with increasing dietary grain in a goat model. *Environmental Microbiology* 18, 525–541.
- Marchesi JR and Ravel J 2015. The vocabulary of microbiome research: a proposal. *Microbiome* 3, 31.
- Martinez-Fernandez G, Denman SE, Cheung J and McSweeney CS 2017. Phloroglucinol degradation in the rumen promotes the capture of excess hydrogen generated from methanogenesis inhibition. *Frontiers in Microbiology* 8, 1871.
- Martinez-Fernandez G, Denman SE, Yang C, Cheung J, Mitsumori M and McSweeney CS 2016. Methane inhibition alters the microbial community, hydrogen flow, and fermentation response in the rumen of cattle. *Frontiers in Microbiology* 7, 1122.
- McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R and Hugenholtz P 2012. An improved greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal* 6, 610–618.
- McMurdie PJ and Holmes S 2014. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Computational Biology* 10, e1003531.
- Mitsumori M, Matsui H, Tajima K, Shinkai T, Takenaka A, Denman SE and McSweeney CS 2014. Effect of bromochloromethane and fumarate on phylogenetic diversity of the formyltetrahydrofolate synthetase gene in bovine rumen. *Animal Science Journal* 85, 25–31.
- Morgavi DP, Rathahao-Paris E, Popova M, Bocard J, Nielsen KF and Boudra H 2015. Rumen microbial communities influence metabolic phenotypes in lambs. *Frontiers in Microbiology* 6, 1060.
- Myer PR, Smith TPL, Wells JE, Kuehn LA and Freetly HC 2015. Rumen microbiome from steers differing in feed efficiency. *PLoS One* 10, e0129174.
- Noel SJ, Attwood GT, Rakonjac J, Moon CD, Waghorn GC and Janssen PH 2017. Seasonal changes in the digesta-adherent rumen bacterial communities of dairy cattle grazing pasture. *PLoS One* 12, e0173819.
- O'Callaghan TF, Vazquez-Fresno R, Serra-Cayuela A, Dong E, Mandal R, Hennessy D, McAuliffe S, Dillon P, Wishart DS, Stanton C and Ross RP 2018. Pasture feeding changes the bovine rumen and milk metabolome. *Metabolites* 8.
- Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O and Vonstein V 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research* 33, 5691–5702.
- Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A and Hugenholtz P 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology* nbt.4229. Retrieved from <https://www.nature.com/articles/nbt.4229.pdf>.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P and Tyson GW 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* 25, 1043–1055.
- Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P and Tyson GW 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology* 2, 1533–1542.
- Patil KR, Haider P, Pope PB, Turnbaugh PJ, Morrison M, Scheffer T and McHardy AC 2011. Taxonomic metagenome sequence assignment with structured output models. *Nature Methods* 8, 191–192.
- Pearson K 1897. Mathematical contributions to the theory of evolution.—on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London* 60, 489–498.
- Peng Y, Leung HC, Yiu SM and Chin FY 2011. Meta-IDBA: a de Novo assembler for metagenomic data. *Bioinformatics* 27, i94–i101.
- Peng Y, Leung HC, Yiu SM and Chin FY 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428.
- Petri RM, Schwaiger T, Penner GB, Beauchemin KA, Forster RJ, McKinnon JJ and McAllister TA 2013. Characterization of the core rumen microbiome in cattle during transition from forage to concentrate as well as during and after an acidotic challenge. *PLoS One* 8, e83424.
- Petri RM, Vahmani P, Yang HE, Dugan MER and McAllister TA 2018. Changes in rumen microbial profiles and subcutaneous fat composition when feeding extruded flaxseed mixed with or before hay. *Frontiers in Microbiology* 9, 1055.
- Petritz BA and Franco OL 2017. Metaproteomics as a complementary approach to gut microbiota in health and disease. *Frontiers in Chemistry* 5, 4.
- Pitta DW, Indugu N, Kumar S, Vecchiarelli B, Sinha R, Baker LD, Bhukya B and Ferguson JD 2016. Metagenomic assessment of the functional potential of the rumen microbiome in Holstein dairy cows. *Anaerobe* 38, 50–60.
- Pope PB, Mackenzie AK, Gregor I, Smith W, Sundset MA, McHardy AC, Morrison M and Eijsink VG 2012. Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. *PLoS One* 7, e38571.
- Poulsen M, Schwab C, Jensen BB, Engberg RM, Spang A, Canibe N, Hojberg O, Milinovich G, Fragner L, Schleper C, Weckwerth W, Lund P, Schramm A and Ulrich T 2013. Methylophilic methanogenic thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nature Communications* 4, 1428.
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J and Glockner FO 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* 35, 7188–7196.
- Qi M, Wang P, O'Toole N, Barboza PS, Ungerfeld E, Leigh MB, Selinger LB, Butler G, Tsang A, McAllister TA and Forster RJ 2011. Snapshot of the eukaryotic gene expression in muskoxen rumen—a metatranscriptomic approach. *PLoS One* 6, e20521.
- Qian C and Hettich RL 2017. Optimized extraction method to remove humic acid interferences from soil samples prior to microbial proteome measurements. *Journal of Proteome Research* 16, 2537–2546.
- Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF and Sloan WT 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods* 6, 639–641.
- Robinson MD, McCarthy DJ and Smyth GK 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Roehe R, Dewhurst RJ, Duthie C-A, Rooke JA, McKain N, Ross DW, Hyslop JJ, Waterhouse A, Freeman TC, Watson M and Wallace RJ 2016. 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 Genetics* 12, e1005846.
- Rooke JA, Wallace RJ, Duthie CA, McKain N, de Souza SM, Hyslop JJ, Ross DW, Waterhouse T and Roehe R 2014. Hydrogen and methane emissions from beef cattle and their rumen microbial community vary with diet, time after feeding and genotype. *British Journal of Nutrition* 112, 398–407.
- Saleem F, Ametaj BN, Bouatra S, Mandal R, Zebeli Q, Dunn SM and Wishart DS 2012. A metabolomics approach to uncover the effects of grain diets on rumen health in dairy cows. *Journal of Dairy Science* 95, 6606–6623.

- Saleem F, Bouatra S, Guo AC, Psychogios N, Mandal R, Dunn SM, Ametaj BN and Wishart DS 2013. The bovine ruminal fluid metabolome. *Metabolomics* 9, 360–378.
- Sarkar N 1997. Polyadenylation of mRNA in prokaryotes. *Annual Review of Biochemistry* 66, 173–197.
- Saro C, Hohenester UM, Bernard M, Lagrée M, Martin C, Doreau M, Boudra H, Popova M and Morgavi DP 2018. Effectiveness of interventions to modulate the rumen microbiota composition and function in pre-ruminant and ruminant lambs. *Frontiers in Microbiology* 9, 1273.
- Seedorf H, Kittelmann S, Henderson G and Janssen PH 2014. RIM-DB: a taxonomic framework for community structure analysis of methanogenic archaea from the rumen and other intestinal environments. *PeerJ* 2, e494.
- Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, Eloe-Fadrosch EA, Pavlopoulos GA, Hadjithomas M, Varghese NJ, Paez-Espino D, Hungate project c, Perry R, Henderson G, Creevey CJ, Terrapon N, Lapebie P, Drula E, Lombard V, Rubin E, Kyrpidis NC, Henrissat B, Woyke T, Ivanova NN and Kelly WJ 2018. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nature Biotechnology* 36, 359–367.
- Shabat SKB, Sasson G, Doron-Faigenboim A, Durman T, Yaacoby S, Berg Miller ME, White BA, Shterzer N and Mizrahi I 2016. Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *The ISME Journal* 10, 2958–2972.
- Shi W, Moon CD, Leahy SC, Kang D, Froula J, Kittelmann S, Fan C, Deutsch S, Gagic D, Seedorf H, Kelly WJ, Atua R, Sang C, Soni P, Li D, Pinares-Patiño CS, McEwan JC, Janssen PH, Chen F, Visel A, Wang Z, Attwood GT and Rubin EM 2014. Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Research* 24, 1517–1525.
- Shinkai T, Mitsumori M, Sofyan A, Kanamori H, Sasaki H, Katayose Y and Takenaka A 2016. Comprehensive detection of bacterial carbohydrate-active enzyme coding genes expressed in cow rumen. *Animal Science Journal* 87, 1363–1370.
- Silverman JD, Washburne AD, Mukherjee S and David LA 2017. A phylogenetic transform enhances analysis of compositional microbiota data. *eLife* 6, e21887.
- Snelling TJ and Wallace RJ 2017. The rumen microbial metaproteome as revealed by SDS-PAGE. *BMC Microbiology* 17, 9.
- Solden LM, Hoyt DW, Collins WB, Plank JE, Daly RA, Hildebrand E, Beavers TJ, Wolfe R, Nicora CD, Purvine SO, Carstensen M, Lipton MS, Spalinger DE, Firkins JL, Wolfe BA and Wrighton KC 2017. New roles in hemicellulosic sugar fermentation for the uncultivated Bacteroidetes family BS11. *The ISME Journal* 11, 691–703.
- Solomon KV, Haitjema CH, Henske JK, Gilmore SP, Borges-Rivera D, Lipzen A, Brewer HM, Purvine SO, Wright AT, Theodorou MK, Grigoriev IV, Regev A, Thompson DA and O'Malley MA 2016. Early-branching gut fungi possess a large, comprehensive array of biomass-degrading enzymes. *Science* 351, 1192–1195.
- Stackebrandt E and Goebel BM 1994. Taxonomic note: a place for dna-dna reassociation and 16s rna sequence analysis in the present species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology* 44, 846–849.
- Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, Liachko I, Snelling TJ, Dewhurst RJ, Walker AW, Roehe R and Watson M 2018. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nature Communications* 9, 870.
- Svartstrom O, Alneberg J, Terrapon N, Lombard V, de Bruijn I, Malmsten J, Dalin AM, El Muller E, Shah P, Wilmes P, Henrissat B, Aspeborg H and Andersson AF 2017. Ninety-nine de novo assembled genomes from the moose (*Alces alces*) rumen microbiome provide new insights into microbial plant biomass degradation. *The ISME Journal* 11, 2538–2551.
- Tapio I, Fischer D, Blasco L, Tapio M, Wallace RJ, Bayat AR, Ventto L, Kahala M, Negussie E, Shingfield KJ and Vilkki J 2017b. Taxon abundance, diversity, co-occurrence and network analysis of the ruminal microbiota in response to dietary changes in dairy cows. *PLoS One* 12, e0180260.
- Timmins-Schiffman E, May DH, Mikan M, Riffle M, Frazar C, Harvey HR, Noble WS and Nunn BL 2017. Critical decisions in metaproteomics: achieving high confidence protein annotations in a sea of unknowns. *The ISME Journal* 11, 309–314.
- Tapio I, Shingfield KJ, McKain N, Bonin A, Fischer D, Bayat AR, Vilkki J, Taberlet P, Snelling TJ and Wallace RJ 2016. Oral samples as non-invasive proxies for assessing the composition of the rumen microbial community. *PLoS One* 11, e0151220.
- Tapio I, Snelling TJ, Strozzi F and Wallace RJ 2017a. The ruminal microbiome associated with methane emissions from ruminant livestock. *Journal of Animal Science and Biotechnology* 8, 7.
- Toyoda A, Iio W, Mitsumori M and Minato H 2009. Isolation and identification of cellulose-binding proteins from sheep rumen contents. *Applied and Environmental Microbiology* 75, 1667–1673.
- van der Ark KCH, van Heck RGA, Martins Dos Santos VAP, Belzer C and de Vos WM 2017. More than just a gut feeling: constraint-based genome-scale metabolic models for predicting functions of human intestinal microbes. *Microbiome* 5, 78.
- Větrovský T and Baldrian P 2013. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* 8, e57923.
- Wallace RJ, Rooke JA, McKain N, Duthie C-A, Hyslop JJ, Ross DW, Waterhouse A, Watson M and Roehe R 2015. The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* 16, 839.
- Wang Q, Garrity GM, Tiedje JM and Cole JR 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied Environmental Microbiology* 73, 5261–5267.
- Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, Lozupone C, Zaneveld JR, Vázquez-Baeza Y, Birmingham A, Hyde ER and Knight R 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5, 27.
- Wetzels SU, Mann E, Pourazad P, Qumar M, Piniore B, Metzler-Zebeli BU, Wagner M, Schmitz-Esser S and Zebeli Q 2017. Epimural bacterial community structure in the rumen of Holstein cows with different responses to a long-term subacute ruminal acidosis diet challenge. *Journal of Dairy Science* 100, 1829–1844.
- Wilkinson TJ, Huws SA, Edwards JE, Kingston-Smith AH, Siu-Ting K, Hughes M, Rubino F, Friedersdorff M and Creevey CJ 2018. CowPI: a rumen microbiome focussed version of the PICRUSt functional inference software. *Frontiers in Microbiology* 9, 1095.
- Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, Vaughan R, Hunter C, Park J, Morrison N, Rocca-Serra P, Sterk P, Arumugam M, Bailey M, Baumgartner L, Birren BW, Blaser MJ, Bonazzi V, Booth T, Bork P, Bushman FD, Buttigieg PL, PSG Chain, Charlson E, Costello EK, Huot-Creasy H, Dawyndt P, DeSantis T, Fierer N, Fuhrman JA, Gallery RE, Gevers D, Gibbs RA, Gil IS, Gonzalez A, Gordon JL, Guralnick R, Hankeln W, Highlander S, Hugenholtz P, Jansson J, Kau AL, Kelley ST, Kennedy J, Knights D, Koren O, Kuczynski J, Kyrpidis N, Larsen R, Lauber CL, Legg T, Ley RE, Lozupone CA, Ludwig W, Lyons D, Maguire E, Methé BA, Meyer F, Muegge B, Nakielny S, Nelson KE, Nemergut D, Neufeld JD, Newbold LK, Oliver AE, Pace NR, Palanisamy G, Peplies J, Petrosino J, Proctor L, Pruesse E, Quast C, Raes J, Ratnasingham S, Ravel J, Relman DA, Assunta-Sansone S, Schloss PD, Schriml L, Sinha R, Smith MI, Sodergren E, Spor A, Stombaugh J, Tiedje JM, Ward DV, Weinstock GM, Wendel D, White O, Whiteley A, Wilke A, Wortman JR, Yatsunenko T and Glöckner FO 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nature Biotechnology* 29, 415–420.
- Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W and Glöckner FO 2014. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Research* 42, D643–D648.
- Zarecki R, Oberhardt MA, Reshef L, Gophna U and Ruppin E 2014. A novel nutritional predictor links microbial fastidiousness with lowered ubiquity, growth rate, and cooperativeness. *PLoS Computational Biology* 10, e1003726.
- Zhang J, Shi H, Wang Y, Li S, Cao Z, Ji S, He Y and Zhang H 2017a. Effect of dietary forage to concentrate ratios on dynamic profile changes and interactions of ruminal microbiota and metabolites in holstein heifers. *Frontiers in Microbiology* 8, 2206.
- Zhang R, Ye H, Liu J and Mao S 2017b. High-grain diets altered rumen fermentation and epithelial bacterial community and resulted in rumen epithelial injuries of goats. *Applied Microbiology and Biotechnology* 101, 6981–6992.
- Zhao S, Zhao J, Bu D, Sun P, Wang J and Dong Z 2014. Metabolomics analysis reveals large effect of roughage types on rumen microbial metabolic profile in dairy cows. *Letters in Applied Microbiology* 59, 79–85.